Mapping forest structural heterogeneity of tropical montane forest remnants from airborne laser scanning and Landsat time series

Adhikari, Hari; Valbuena, Ruben; Pellikka, Petri K.E.; Heiskanen, Janne

Ecological Indicators

DOI:
https://doi.org/10.1016/j.ecolind.2019.105739

Published: 01/01/2020

Publisher's PDF, also known as Version of record

Ecological Indicators 108 (2020) 105739

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Mapping forest structural heterogeneity of tropical montane forest remnants from airborne laser scanning and Landsat time series

Hari Adhikari[a,b,*], Ruben Valbuena[c,d], Petri K.E. Pellikka[a,b], Janne Heiskanen[a,b]

[a] Earth Change Observation Laboratory, Department of Geosciences and Geography, University of Helsinki, P.O. Box 68, FI-00014 Helsinki, Finland
[b] Institute for Atmospheric and Earth System Research, Faculty of Science, University of Helsinki, P.O. Box 68, FI-00014 Helsinki, Finland
[c] Bangor University, School of Natural Sciences, Thoday Building, LL57 2UW Bangor, United Kingdom
[d] University of Eastern Finland, School of Forest Sciences, P.O. Box 111, FI-80100 Joensuu, Finland

ARTICLE INFO

Keywords:
Forest structure
Gini coefficient
Spectral-temporal metrics
LiDAR
Africa

ABSTRACT

Tropical montane forests are important reservoirs of carbon and biodiversity and have a central role in the hydrological cycle. They are, however, very fragmented and degraded, leaving isolated remnants across the landscape. These montane forest remnants have considerable differences in forest structure, depending on factors such as tree species composition and degree of forest degradation. Our objectives were (1) to analyse the reliability of airborne laser scanning (ALS) in modelling forest structural heterogeneity, as described by the Gini coefficient (GC) of tree size inequality; (2) to determine whether models are improved by including tree species-sensitive spectral-temporal metrics from the Landsat time series (LTS); and (3) to evaluate differences between three forest remnants and different forest types using the resulting maps of predicted GC. The study area was situated in Taita Hills, Kenya, where indigenous montane forests have been partly replaced by single-species plantations. The data included field measurements from 85 sample plots and two ALS data sets with different pulse densities (9.6 and 3.1 pulses m^{-2}). GC was modeled using beta regression. We found that GC was predicted more accurately by the ALS data set with a higher point density (a cross-validated relative root mean squared error ($rRMSE_{CV}$) 13.9%) compared to ALS data set with lower point density ($rRMSE_{CV}$ 15.1%). Furthermore, important synergies exist between ALS and LTS metrics. When combining ALS and LTS metrics, $rRMSE_{CV}$ was improved to 12.5% and 13.0%, respectively. Therefore, if the LTS metrics are included in models, ALS data with lower pulse density are sufficient to yield similar accuracy to more expensive, higher pulse density data acquired from the lower altitude. In Ngangao and Yale, forest canopy has multiple layers of variable tree sizes, whereas elfin forests in Vuria are of more equal tree size, and the GC value ranges of the indigenous forests are 0.42–0.71, 0.20–0.74, and 0.17–0.76, respectively. The single-species plantations of cypress and pine showed lower values of GC than indigenous forests located in the same remnants in Yale, whereas Eucalyptus plantations showed GC values more similar to the indigenous forests. These results show the usefulness of GC maps for identifying and separating forest types as well as for assessing their distinctive ecologies.

1. Introduction

The Eastern Arc Mountains (EAM) are a chain of crystalline Precambrian basement mountains, stretching from southern Kenya to eastern Tanzania (Burgess et al., 2007). The tropical montane forests in the EAM are important reservoirs of carbon and biodiversity (Adhikari et al., 2017; Burgess et al., 2007; Lovett and Wasser 1993). Furthermore, the montane forests capture moisture and store precipitation on the hilltops, hence having a central role in the hydrological cycle (Pellikka et al., 2009). However, montane forests of EAM, once dense and continuous stands of indigenous forest, are now threatened and fragmented due to land-use change and conversion to cropland and agroforestry, which leaves a patchy landscape of forest remnants (Pellikka et al., 2013). Depending on such factors as tree species composition and degree of forest degradation, the extant montane forest remnants have considerable differences in forest structure with effects on carbon stocks (Pellikka et al., 2018; Omoro et al., 2013), biodiversity (Thijs et al., 2014), and forest function (Pfeifer et al., 2018).

Vertical structural heterogeneity describes the size variability of plant assemblages directly competing for resources at a local scale, thus

* Corresponding author at: Earth Change Observation Laboratory, Department of Geosciences and Geography, University of Helsinki, P.O. Box 68, FI-00014 Helsinki, Finland.
E-mail address: hari.adhikari@helsinki.fi (H. Adhikari).

https://doi.org/10.1016/j.ecolind.2019.105739
Received 11 June 2019; Received in revised form 12 August 2019; Accepted 13 September 2019
Available online 23 September 2019

reflecting characteristics of the ecosystem's complexity such as conditions of competitive dominance of resource depletion or pre-emption in a given ecosystem (Weiner, 1990), or habitat disturbance history (Valbuena et al., 2016). The structural complexity of ecosystems provides habitat quality for fauna, affecting species diversity (Willson, 1974; Erdelen, 1984). Therefore, changes in morphological traits describing the vertical complexity of the plant ecosystem may be key to determining a threat to the overall biodiversity (MacArthur and MacArthur, 1961; Camargo and Kapos, 2009).

Airborne laser scanning (ALS; a.k.a. airborne LIDAR) provides a method for high–resolution mapping of forest three-dimensional structure via direct measurements of forest canopy height and canopy cover and through predictive modeling of forest structural attributes. When using ALS, some authors express proportions of leaf area density (LAD) along vertical strata, e.g. overstorey versus understorey (Miura and Jones, 2010; Lesak et al., 2011; Hill et al., 2014; Almeida et al., 2019). Understorey density has been found to be valuable in determining the dispersion of key animal species (Clawges et al., 2008; Jung et al., 2012; Zellweger et al., 2017). Indicators to concisely summarize plant height distributions are fundamentally based on either the entropy (structural diversity) or the equitability (size variance) of the plant size distribution (Valbuena et al., 2012).

Many studies have employed ALS derivations of McArthur and McArthur's (1961) foliage height diversity (FHD) as a morphological trait describing the complexity of the plant community (Schneider et al., 2017), showing its relationship to species distributions and biodiversity (Vierling et al., 2008; Bergen et al., 2009). On the other hand, the most widespread alternatives to express size variance are the standard deviation (McRoberts et al., 2008, Coops et al., 2016) and the Gini coefficient (GC) (Valbuena et al., 2013b, 2017c; Dalponte et al., 2018; Mononen et al., 2018; Zhang et al., 2019; Erfanifard et al., 2019) of ALS heights. Descriptors of tree size variability are also good proxies for biodiversity. Goetz et al. (2007), for instance, found that the total species richness of the community was related to a ALS distribution ratio. GC is calculated based on the basal area of the individual trees, which represent the tree size inequality. Compared to other measures that can be used to evaluate tree size variability, such as the standard deviation, the GC brings the advantage of being independent from the mean tree size, and being bounded by meaningful theoretical values such as such GC = 0 representing total equality and GC = 1 maximum inequality, or GC = 0.5 representing maximum entropy (Valbuena et al., 2012). GC has been used as a proxy to measure forest structural diversity (Pach and Podlaski, 2015) and biodiversity (Nölte et al., 2018). Many studies have examined modeling GC using ALS data for assessing boreal forest structure (Goodwin et al., 2006; Valbuena et al., 2014; Valbuena et al., 2016), but the method has not previously been assessed in tropical montane forest landscapes.

Although ALS can provide highly accurate forest structure estimates, it is possible that the addition of independent phenological data that correlate with forest biophysical properties, e.g. multispectral data, can further improve its ability to predict GC. Incorporating simultaneously acquired ALS and high–resolution multispectral imagery improved species classification at tree level (Ørka et al., 2012). There has been some research on synergistic use of ALS and high spatial resolution spectral metrics from aerial imagery to predict forest structural attributes (e.g. forest canopy cover, canopy height, GC, and density) (Ahmed et al., 2014, Hudak et al., 2002, Manzanera et al., 2016, Valbuena et al., 2017b), but use is limited due to the costs involved in collecting high-resolution data.

The open (free) data policy of the Landsat satellite mission has revolutionized the use of Landsat data (Zhu et al., 2019). Multispectral Landsat time-series data have been applied successfully for mapping tree height distributions in sub-Saharan Africa (Hansen et al., 2016) and above-ground biomass in tropical forests (Phua et al., 2017). However, the medium spatial resolution data, e.g. Landsat 8 OLI data, have not been tested for modeling GC to date in the Afromontane forests. Furthermore, there are no studies on combining ALS metrics and spectral-temporal metrics from the Landsat time series (LTS) for GC mapping.

In this study, our objective was to characterize the variation of forest structure in three Afromontane forest remnants in the Taita Hills. More specifically, we aimed (1) to analyse how well tree size inequality, described by GC, can be modeled by two airborne laser scanning data sets acquired from different altitudes using two different sensors; (2) to determine whether models are improved by including tree species-sensitive LTS; (3) to generate GC maps for the forest remnants and evaluate differences between the three forest remnants and indigenous and exotic plantation forests.

2. Material and methods

2.1. Study area

The Taita Hills montane forests in south-eastern Kenya are part of the EAM. The semi-arid plains isolate Taita Hills from the other hills, e.g. Pare and Usambara Mountains in Tanzania and Mount Kasigau situated 50 km to the southeast of Taita Hills in Kenya. An average elevation at the lower plains is 700 m above sea level (a.s.l.) and at the hills 1500 m a.s.l., with the highest hilltop (Vuria) reaching 2208 m a.s.l. (Pellikka et al., 2018).

The landscape is fragmented due to the conversion of forested areas to agricultural lands, while at the same time exotic plantations are established on, for example, erosion-prone areas. The forest is currently limited to small fragments of indigenous and plantation forests between 50 and 200 ha in size. The indigenous cloud forest fragments are restricted to above 1400 m altitude on the southeastern slopes and above 1700 m on the northwestern slopes, in the areas receiving over 900 mm of precipitation annually (Pellikka et al., 2009). In our study, we focused on three forest remnants: Ngangao, Yale, and Vuria (Fig. 1), located on the mountain ridges with steep slopes at an altitude range of 1700–1952 m, 1750–2104 m, and 1655–2208 m, respectively (Fig. 2).

The study area has a bimodal rainfall annual regime, with long rains from March to June and short rains between October and December. Annual rainfall in Taita Hills is according to Erdogan et al. (2011) between 1100 and 1400 mm, while in the lowlands it is between 400 and 600 mm. The upper montane cloud forests, like Vuria, receive abundant moisture from the low-lying clouds, fog, and moisture-laden southeast trade winds originating from the Indian Ocean (Räsänen et al., 2018). These forests are typically single-layered and covered by epiphytic mosses and lichens. In contrast, the lower montane forests, like Ngangao, are drier and taller with multi-layered tree canopy (Stam et al., 2017). A large part of Yale mountain is barren rock, heathland, or rocks covered by *Acacia mearnsii* (hereafter wattle tree), while the steep slopes are covered by *Eucalyptus* spp. (hereafter eucalyptus) and *Cupressus lusitanica* (hereafter cypress) (Pellikka et al., 2009). Each of the three forest fragments consists of indigenous montane forest parts with several planted exotic trees (Fig. 2).

In Taita Hills, the total forest cover remained virtually unchanged between 1955 and 2004 (Pellikka et al., 2009). However, 50% of the indigenous montane forest was cleared during that period due to climatic and edaphic conditions favorable for agriculture, and because new exotic plantation forests were established (Pellikka et al., 2009). In 1955–2004 for Ngangao and 1955–1994 for Yale, the change of non-forested area to the indigenous forest was 12.0 ha and 11.1 ha, and to exotic forest 81.5 ha and 94.2 ha, respectively (Pellikka et al., 2009). The indigenous forests of Taita Hills are rich in species; for example, in Ngangao 73 woody tree species were identified in the study plots by Mbuthia (2003) and 52 tree species in the plots by Schäfer et al. (2016). The most common indigenous tree species in the montane forests include *Albizia gummifera*, *Macaranga capensis,* and *Maesa lanceolate.* The plantation forests established between the 1950s and 1970s include stands of eucalyptus, cypress, and pine (*Pinus caribea, Pinus elliottii*, *and*

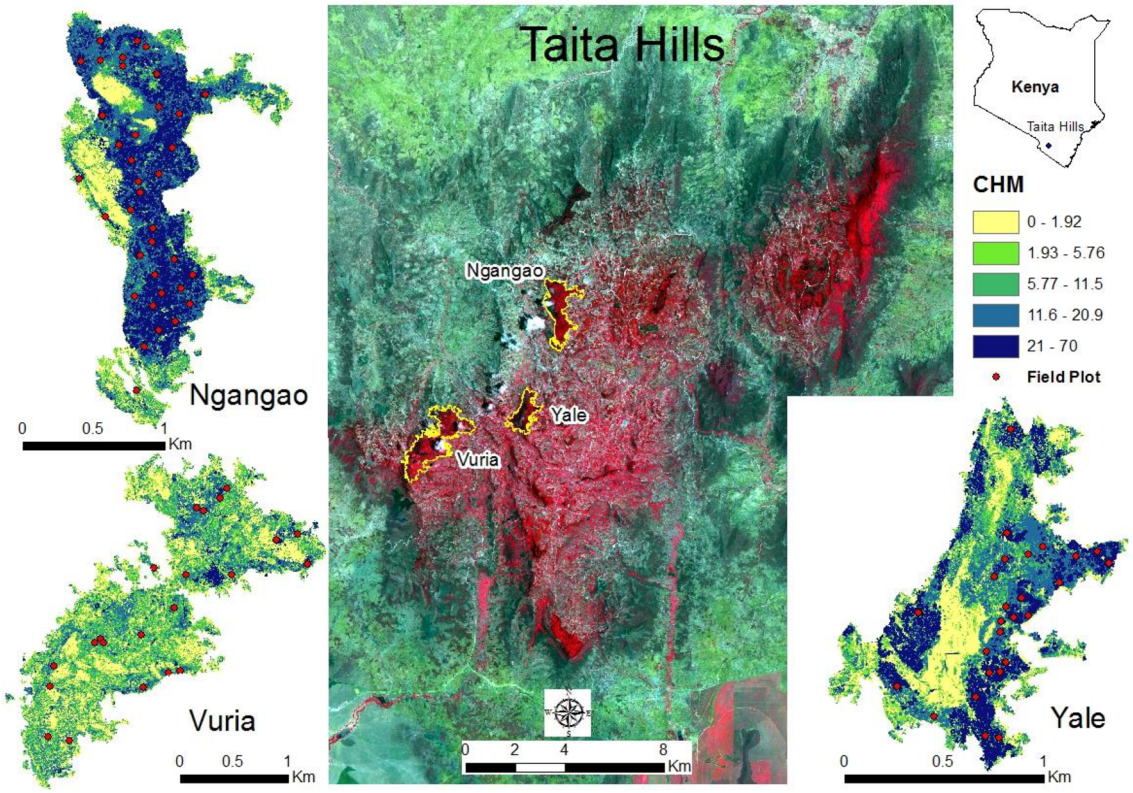

Fig. 1. Location of Taita Hills and the three forest remnants studied there (Ngangao, Yale, and Vuria) with a Sentinel-2A MSI satellite image from 8 October 2016. The canopy height model (CHM) for the forests is shown in the inset maps.

Pinus patula), and between the 1970s and 1980s, *Grevillea robusta* (hereafter grevillea) and *Maesopsis eminii* (Adhikari et al., 2017; Pellikka et al., 2009). Exotic trees were planted for soil erosion protection and production of wood and building material (Pellikka et al., 2009).

Other important canopy, subcanopy, and understorey tree species are *Craibia zimmermannii, Cola greenwayi, Newtonia buchananii, Tabernaemontana stapfiana, Strombosia scheffleri, Syzygium sclerophyllum, Ochna holstii,* and *Macaranga conglomerate* in the moist montane forest (Ngangao) and *Prunus africana* and *Philippia pallidiflora* in high altitude and upper montane forests (Yale and Vuria) (Aerts et al., 2011). The high dissimilarity in woody plant communities may be an effect of historical or recent isolation (Aerts et al., 2011), which cannot be confirmed due to the absence of detailed historical forest maps. Selective logging of the most valuable tree species, such as *Ocotea usambarensis*, has taken place in forests, evidenced in Ngangao by, for instance, hundreds of sawpits (Boström 2010), which have reduced the multi-layer characteristics of the forests.

2.2. Field data

The field measurements from 85 circular 0.1 ha sample plots (radius 17.84 m) collected between 2013 and 2015 were used for forest structural diversity modeling (Fig. 1). This plot size secures a robust and unbiased estimation of GC, as described in Adnan et al. (2017). The sampling of forest plot was guided by canopy height model (CHM) based on ALS data set (see Table 1 for more details) and visible to near-infrared imaging spectroscopy data (AisaEAGLE) acquired in 2013 (Piiroinen et al., 2018). The number of plots in Ngangao, Yale, and Vuria were 37, 27, and 21, respectively. Sample plot centers were positioned using Trimble GeoXH GNSS receiver, and the differential correction were made using a GNSS base station set up at the Taita Research Station, Wundanyi. For all tree stems with a diameter (D) at breast height (1.3 m) $>$ 10 cm, D was measured (Adhikari et al., 2017). Most of the tree species were identified by a local para-taxonomist and a field guide.

GC based on Lorenz curve quantifies tree size inequality among trees within the forest (see also Valbuena et al., 2013a) and was used as a forest structure indicator. Lorenz curves and GCs for each sample plot were calculated based on basal areas (area occupied by a given D) of individual trees (Valbuena et al., 2016). GC is calculated as the area between the line of equality and Lorenz curve divided by the total area below the line of equality. The GC ranges from 0 to 1, where 0 represents perfect equality (all trees of equal size) and 1 represents perfect inequality (few trees have the largest share) (Nölte et al., 2018). According to Valbuena et al. (2012), GC $<$ 0.5, GC = 0.5, and GC $>$ 0.5 represent even-sized forest, irregular forest, and bimodal diameter distributions, respectively.

2.3. Airborne laser scanning data

ALS data were collected using two sensors in three different years (Table 1). ALS1 and ALS2 are different in flight altitude, pulse density, and sensor. Both data sets were preprocessed and delivered as georeferenced point clouds in the UTM/WGS84 coordinate system with ellipsoidal heights by the data vendors Topscan Gmbh and Ramani Geosystems (Heiskanen et al., 2015). Returns were classified as ground returns and used to produce a digital terrain model (DTM) with a 2 m cell grid using LAStools software (Isenburg, 2014). The ALS point cloud elevations were normalized to heights from ground level using DTM from the corresponding campaign. Noisy returns (e.g. high points) and returns from electric lines were removed manually.

ALS returns corresponding to the field plots were extracted based on plot center coordinates and fixed plot radii. ALS metrics for predictive models of forest structure attributes were calculated for each field plot using FUSION software (McGaughey, 2016). A 3-m height threshold

Fig. 2. (a) Western slopes of Ngangao forest on the left, the peak of Yale in the middle, and the highest mountain Vuria on the right within an agricultural landscape mosaic of Taita Hills. (b) The three peaks of Vuria, representing mixed indigenous forest in the highest peak and eucalyptus forest on the right. (c) Stands of pine, eucalyptus, cypress, and indigenous forests on the eastern slope of Yale. (d) Indigenous lower montane forest on the northern part of Ngangao in front, and pine and cypress stands behind. (e) Degraded indigenous forest in the middle peak of Vuria (Photos: P. Pellikka).

Table 1
Characteristics of airborne laser scanner (ALS) data.

Parameter	ALS1	ALS2
Date of acquisition	2013	2014 and 2015
Sensor	Optech ALTM 3100	Leica ALS60
Mean flying height above ground (m)	760	1460
Pulse rate (kHz)	100	58
Scan rate (Hz)	36	66
Maximum scan angle (degrees)	± 16	± 16
Mean pulse density (pulses m^{-2})	9.6	3.1
Mean return density (returns m^{-2})	11.4	3.4
Maximum number of returns per pulse	4	4
Mean footprint diameter (cm)	23	32

was used to separate understorey and ground returns from canopy returns (Adhikari et al., 2017; Heiskanen et al., 2019). Canopy cover metrics were calculated from all returns (single, first, last), while the height metrics were calculated from the first and last returns (Appendix Table A1) (Gorgens et al., 2017).

2.4. Landsat time series

Landsat 8 Operational Land Imager (OLI) (Landsat Collection 1 Level-2 surface reflectance) images between 14 April 2013 and 20 April 2015 (two complete years – 42 images in total) were downloaded from the United States Geological Survey (USGS) Earth Explorer platform. Spectral-temporal metrics were computed based on Landsat bands 2–7 [Blue (452–512 nm (nm)), Green (533–590 nm), Red (636–673 nm), Near-infrared (NIR) (851–879 nm), Shortwave Infrared 1 (SWIR 1) (1566–1651 nm), and Shortwave Infrared 2 (SWIR 2) (2107–2294 nm)] and vegetation indices, including Normalized Difference Vegetation Index (NDVI), Reduced Simple Ratio (RSR), and Tasseled Cap (TC) (Brightness, Greenness and Wetness), using the statistical distribution of annual band and vegetation index values (Adhikari et al., 2015, 2016; Heiskanen et al., 2019) (Appendix Table A1). The Japan Aerospace Exploration Agency (JAXA) digital elevation model (DEM) was used to carry out topographic normalization of Landsat images (JAXA, 2015). The LTS included several percentile values (10%, 25%, 50%, 75%, and 90%), trimmed means (10% and 25%), inter-percentile range (10–90), and interquartile range (25–75) for all the bands and five vegetation indices (Adhikari et al., 2016; Potapov et al., 2012). The standard deviation of seasonal features was also calculated for 3 × 3 pixel windows centered at each plot. NDVI and RSR are robust against topographic effects (Adhikari et al., 2016). Therefore, no topographic normalization was performed for these indices.

2.5. Statistical modeling and predictor variable selection

Fig. 3 presents our workflow. Data management and statistical analysis were done using software “R” (RCoreTeam, 2017) together

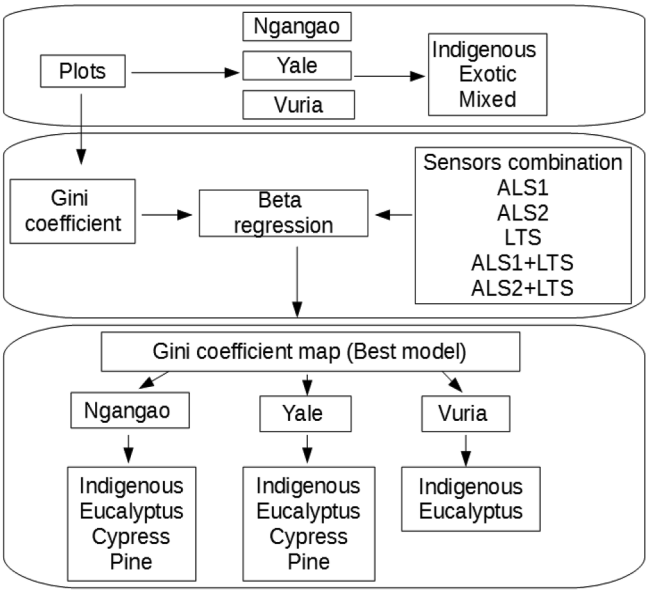

Fig. 3. Workflow employed and stratification of the resulting maps.

with the "betareg" package (Cribari-Neto and Zeileis, 2010) and "leaps" package (Lumley, 2017). Beta regression was used for modeling GC because it is a modeling alternative well suited for these types of response variables ranging between 0 and 1 (Valbuena et al., 2013b). Furthermore, beta regression allows regression in conditions when the distributions in the response variables other than normal. A two-step predictor variable selection strategy was used due to a large number of predictor variables within each predictor group. Firstly, out of many highly inter-correlated metrics (Spearman correlation coefficient > 0.9), the one that had higher correlation with the variable of interest (here GC) was retained. Secondly, probable best predictors were identified using "regsubset" function in "leaps" package. Experiments on different link functions indicated the best performance of "loglog" link function and it was used in all the models (Cribari-Neto and Zeileis, 2010).

Predictive models were calculated using ALS1 metrics only (ALS1), ALS2 metrics only (ALS2), LTS metrics only (LTS), both ALS1 and LTS metrics combined (ALS1 + LTS), and both ALS2 and LTS metrics combined (ALS2 + LTS). The best model for each combination was identified using the accuracy assessment detailed below.

2.6. Accuracy assessment

All of the GC models were assessed and compared by leave-one-out cross-validation (CV). One field plot (i) was taken out each time to eliminate the influence of that plot during model fitting, and the remaining field plots were used to predict a value of GC for that plot (pre_i^{CV}) using beta regression. Hereafter, the abbreviations/subscripts/superscripts "*CV*" and "*fit*" represent measures calculated using CV and non-cross-validated measures (i.e. from model fit residuals), respectively. The model coefficients were averaged from all the iterations to obtain the final model. pre^{CV} was compared against those measured in the field (obs). The relative mean difference (rMD_{cv}) (Eq. (1)) was used to detect any under- or over-prediction. Prediction precision was evaluated using the cross-validated root mean square error ($RMSE_{CV}$) (Eq. (2)) and relative RMSE ($rRMSE_{CV}$) (Eq. (3)). The degree of agreement was evaluated using the coefficient of determination (R^2) (Eqs. (4) and (5)) (Valbuena et al., 2019). Furthermore, Piñeiro et al., (2008) hypothesis tests were used to evaluate whether the observed and predicted values follow the 1:1 correspondence line, based on the null hypothesis that the intercept (α) and slope (β) of the linear regression between the observed and predicted values are H0: $\alpha = 0$ and H0: $\beta = 1$ (Eq. (6)), respectively (Valbuena et al., 2017a). If the null hypothesis for the slope is rejected, the predictions are not consistent with observed values, while if the slope hypothesis is accepted and the null hypothesis for the intercept is rejected, then the model is biased (Piñeiro et al., 2008). If both (slope and intercept) null hypotheses are rejected due to significant $\beta < 1$, then we detect an averaging effect caused by over-fitting (Valbuena et al., 2017a). Finally, to avoid the degree of overfitting to the sample employed in the model fitting, the sum of squares ratio (SSR) (Eq. (7)) was used to limit the models to SSR ≤ 1.1 (Valbuena et al., 2017a; Valbuena et al., 2017b). SSR is the ratio between the square root of the residual sums of squares attained in the cross-validation (SS^{CV}) (Eq. (8)) and that using the whole data set (SS^{fit}) (Eq. (9)).

$$rMD_{cv} = 100 \times \frac{(\sum_{i=1}^{n} (pre_i^{CV} - obs_i)/n)}{\overline{obs}} \quad (1)$$

$$RMSE_{CV} = \sqrt{\sum_{i=1}^{n} (pre_i^{CV} - obs_i)^2 \Big/ n} \quad (2)$$

$$rRMSE_{CV} = 100 \times RMSE_{CV}/\overline{obs} \quad (3)$$

$$R^2_{fit} = 1 - \sum_{j=1}^{n} (pre_j^{fit} - obs_j)^2 \Big/ \sum_{j=1}^{n} (obs_j - \overline{obs})^2 \quad (4)$$

$$R^2_{CV} = 1 - \sum_{i=1}^{n} (pre_i^{CV} - obs_i)^2 \Big/ \sum_{i=1}^{n} (obs_i - \overline{obs})^2 \quad (5)$$

$$obs_i = \alpha + \beta \times pre_i^{CV} \quad (6)$$

$$SSR = \sqrt{SS^{CV}}/\sqrt{SS^{fit}} \quad (7)$$

$$SS^{CV} = \sum_{i=1}^{n} (pre_i^{CV} - obs_i)^2 \quad (8)$$

$$SS^{fit} = \sum_{j=1}^{n} (pre_j^{fit} - obs_j)^2 \quad (9)$$

2.7. Forest structure characterization and analysis

The boundaries of the three forest remnants and different forest types within the forest remnants were mapped using a high-resolution false-colour orthomosaic and CHM based on ALS1. The orthomosaic was based on airborne hyperspectral data (visible to NIR bands, 400–1000 nm) acquired using AisaEAGLE sensor (Specim Ltd., Finland) on 3–8 February 2013 (Piiroinen et al., 2018). AisaEAGLE is a pushbroom scanner with an instantaneous field of view of 0.648 mrad and a field of view of 36.04°. The sensor produced 129 bands with an output pixel resolution of one meter. The bandwidth of each band varies between 4.5 and 5.0 nm (Piiroinen et al., 2018). Only three bands (Green (571 nm), Red (693 nm), and NIR (811 nm)) were used for mapping forest boundaries.

First, AisaEAGLE data were segmented based on the spectral information in ArcGIS 10.3 using Segment Mean Shift tool (ESRI, 2015). Spectral data, tree height (CHM), and field information were then used to identify forest boundaries and to clump the segments manually into stands of the same forest type. Furthermore, orienteering maps of Ngangao (scale 1:10 000) (Boström, 2010), a forest stand map of Vuria (scale 1:10 000) (Boström, 2013), and field plots were used for verifying stand boundaries and dominant species. Each polygon was classified as one of the following types: indigenous, eucalyptus, cypress, pine, bushland, woodland, and rock. For this study, only indigenous, eucalyptus, cypress, and pine were used because these occurred in most of the three forest remnants. Finally, we used the best models to predict GC at 30 m × 30 m grid size over Ngangao, Yale, and Vuria forests.

Based on the maps, we analyzed how GC varies between the moist

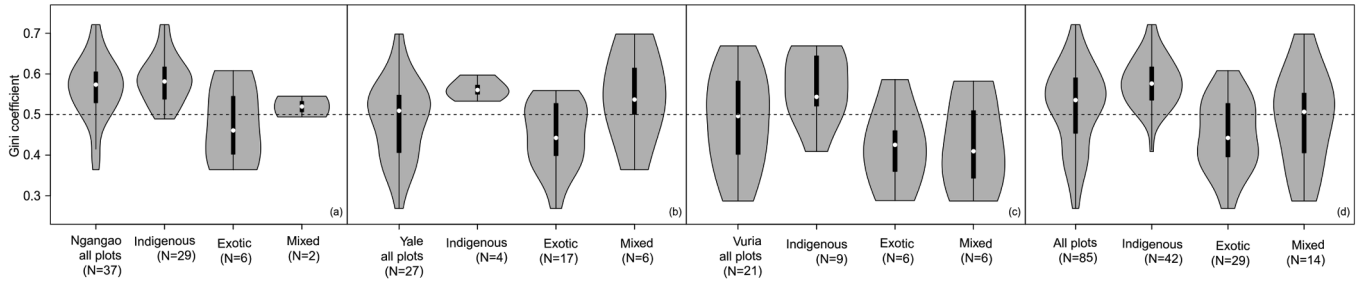

Fig. 4. Distribution of Gini coefficient in each forest remnant and forest type based on the field plots: (a) Ngangao, (b) Yale, (c) Vuria, and (d) all plots pooled together. Indigenous class includes plots dominated by the indigenous species, exotic class includes eucalyptus, cypress, and pine plantations, and mixed class includes both indigenous and exotic species. The white dot represents the median, the thick black bar in the center represents the interquartile range, and the thin black line represents the rest of the distribution (1.5 × interquartile range). Horizontal line represents GC = 0.5.

montane and high altitude montane forests, and between indigenous forests and exotic plantations. Wilcoxon signed-rank test (one sample), Wilcoxon rank–sum tests (also known as Mann–Whitney *U* test), and Kruskal-Wallis test were used to assess whether the median GC in each forest type is above or below the 0.5 thresholds stated in Valbuena et al. (2012), and whether the GC values were significantly different between the forest types compared (R functions "wilcox.test" and "kruskal.test"). The non-parametric tests were used since normality could not be assumed for each forest. In the case of GC based on the maps, more detailed analyses were done for the main exotic plantation species.

3. Results

3.1. Variation of Gini coefficient based on field data

The GC values calculated from the field plots clearly showed per-stratum distributions other than normal (Fig. 4), and for this reason, we hereby report their medians (henceforth denoted as $\widetilde{GC}$) to describe their location, and their interquartile ranges (IQR) as a descriptor for their dispersion. GC < 0.5 represents even-sized forests, GC = 0.5 represents irregular forests, and GC > 0.5 represents bimodal diameter distributions or negative exponential. Overall, tree size heterogeneity in the study area was above the $\widetilde{GC} > 0.5$ threshold in Ngangao (H_0: $\widetilde{GC} \leq 0.5$, H_1: $\widetilde{GC} > 0.5$, $p < 0.001$); however, we failed to prove that for Yale ($\widetilde{GC} = 0.51$, IQR = 0.14) or Vuria ($\widetilde{GC} = 0.50$, IQR = 0.18) when using field plot data only. When a plot had > 70% of its basal area covered by the indigenous species, it was considered pure "indigenous", whereas plots were classified as "exotic" when > 70% were exotic species (eucalyptus, cypress, and pine combined). The rest of the plots were considered "mixed". The $\widetilde{GC}$ and IQR (in parentheses) for the indigenous, exotic, and mixed plots were 0.58 (0.08), 0.46 (0.14), and 0.52 (0.03) in Ngangao, 0.56 (0.02), 0.44 (0.13), and 0.54 (0.11) in Yale, and 0.54 (0.12), 0.43 (0.10), and 0.41 (0.17) in Vuria, respectively.

Results derived from analysing the field data showed that indigenous forests have significantly higher GC than exotic stands (H_0: $\widetilde{GC_{Ind}} \leqslant \widetilde{GC_{Exo}}$, H_1: $\widetilde{GC_{Ind}} > \widetilde{GC_{Exo}}$, p = 0.009 for Ngangao, p = 0.002 for Yale, and p = 0.009 for Vuria) (Fig. 3). This indicates the presence of bimodal diameter distributions in the indigenous forests (H_0: $\widetilde{GC_{Ind}} \leq 0.5$, H_1: $\widetilde{GC_{Ind}} > 0.5$, $p < 0.001$ for Ngangao, p = 0.033 for Yale, and p = 0.025 for Vuria), and exotic plantation forests show even-sized distributions (H_0: $\widetilde{GC_{Exo}} \geq 0.5$, H_1: $\widetilde{GC_{Exo}} < 0.5$, p = 0.025 for Yale, p = 0.057 for Vuria). However, using field plot data only we failed to reject this null hypothesis for Ngangao ($\widetilde{GC} = 0.46$, IQR = 0.142).

Examples of the GC and ALS data are shown for one indigenous forest plot and one exotic plantation forest plot in Fig. 5. The indigenous forest plot has a stratum of smaller trees in DBH range of 10–20 cm along with another stratum of large trees having DBH > 20 cm (Fig. 5a). In these plots, Lorenz curves are shown simultaneously with the diameter distributions. Trees were ordered from the largest to the smallest basal area (left to right, x-axis) and cumulative share of basal area (top to bottom, y-axis) (Fig. 5a and b). In the case of Fig. 5a showing the indigenous forest, its Lorenz curve illustrates that a smaller quantity of big trees (~30%) have a higher proportion of the total basal area (~76%), whilst a larger quantity of small trees (~70%) have a lower proportion of the total basal area (~24%). On the other hand, the plantation forest plot has a more equal DBH distribution with a single vertical stratum since trees were planted at the same time (Fig. 5b). Fig. 5c–f show the point clouds of ALS1 and ALS2 for the same plots. Due to the lower point density in ALS2, the point clouds have fewer echoes reaching the ground and less backscatter from the understorey (Fig. 5d, f).

3.2. Modeling results

Table 2 summarizes the GC modeling results based on ALS metrics only, LTS metrics only, and their combination. In total, 23 regression models were obtained with the number of metrics ranging from one to six depending on the sensor combination (see Appendix Table A2 for selected variables and Table A3 for model parameters). Low values of SSR (SSR < 1.1) indicate no over-fitting problems. For all models, the null hypothesis was rejected (i.e. $\alpha \neq 0$ and $\beta \neq 1$), and the p-values were non-significant, which ensures 1:1 correspondence between the observed and predicted values.

The $rRMSE_{CV}$ for the best models fitted with ALS1 and ALS2 metrics was 13.9% and 15.1%, respectively. The model using only LTS metrics underperformed relative to those using ALS metrics. The coefficient of determination was $R^2_{CV} = 0.28$, and $rRMSE_{CV} = 16.5\%$ was the highest uncertainty of all the alternatives tested (Table 2). The fusion of ALS and LTS metrics improved the models in comparison with using either of them alone. R^2_{CV} was 0.59 and 0.55 for the models combining LTS with ALS1 and ALS2, respectively. The model combining ALS1 and LTS had the lowest $rRMSE_{CV}$, 12.5%, and with ALS2 it was as low as 13.0% (Table 2).

We observed differences according to the pulse density with regard to the ALS metrics, which were selected as the best model predictors (see details in Table A2 in the Appendix). In the case of ALS1, the metrics representing canopy cover, the coefficient of variation of the first returns, L-moment coefficient of variation, the 4th L-moment, and the 5th height percentile of the first and last returns were frequently selected in the models when ALS1 was used, either alone or in combination with LTS metrics. However, in the case of ALS2, in addition to those selected in ALS1, additional ALS metrics included 3rd L-moments kurtosis and L kurtosis height of the first returns, height kurtosis, and L-moment coefficient of variation of the last returns, skewness, and 1th and 99th height percentile for the first return.

The most frequently selected model predictors in case of LTS included, for example, percentile values (10% and 90%), and inter-percentile and interquartile range ('range10_90', and 'range25_75') of the

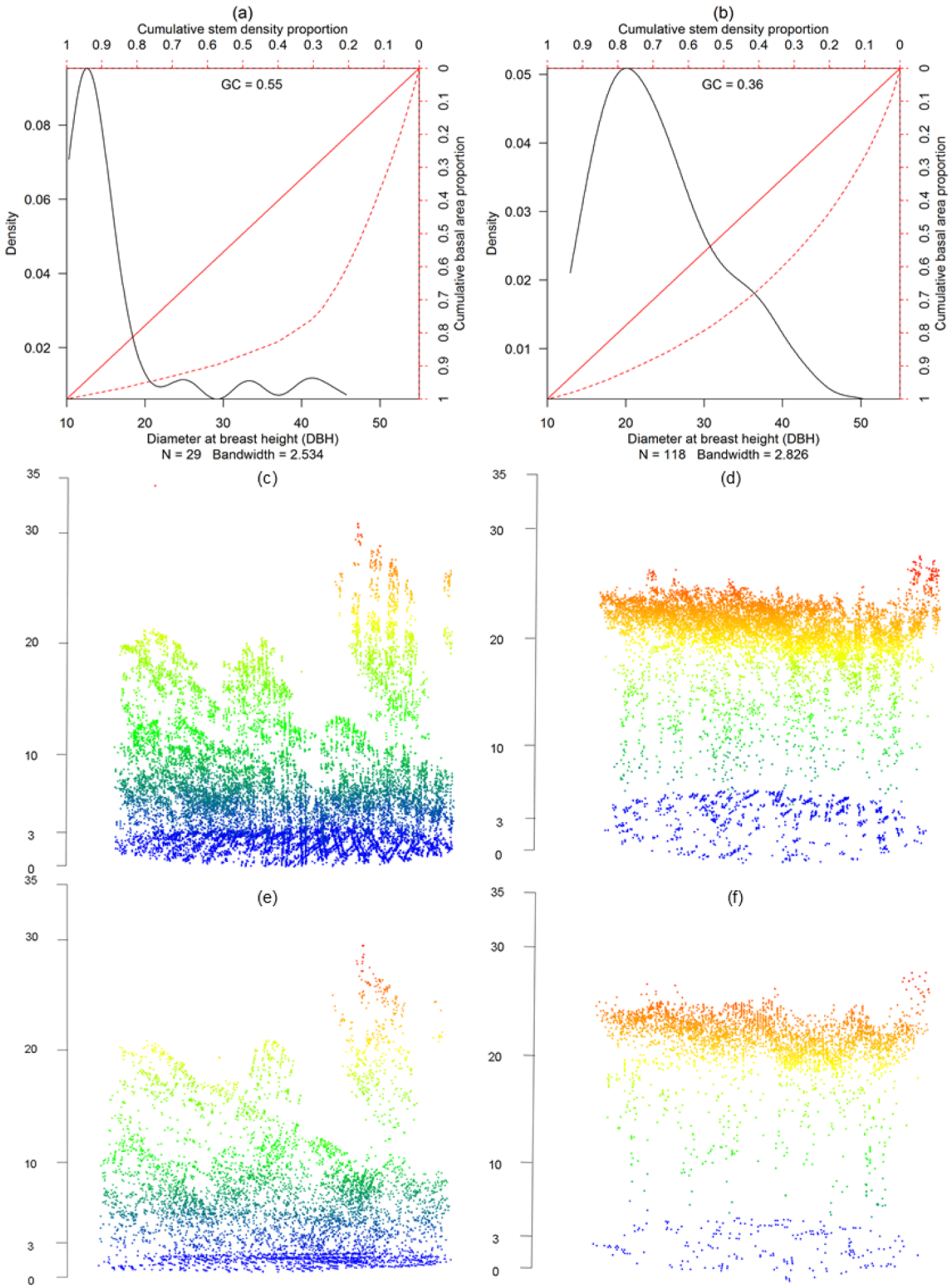

Fig. 5. Diameter at breast height (DBH) kernel density function and Gini coefficient for examples of (a) indigenous and (b) exotic plantation forest (cypress) plots. The straight red line is the line of equality and the dotted line is the Lorenz curve. Minimum DBH was 10 cm. ALS point clouds for the same plots are shown below: ALS1 for indigenous (c) and exotic plots (d), and ALS2 for indigenous (e) and exotic plots (f). N indicates number of trees in these plots.

vegetation indices (NDVI and RSR) and spectral bands (red, NIR, and SWIR). In the case of ALS1 + LTS, high percentile values (90%) for NIR and interquartile range of NDVI and RSR were frequently selected. In the case of ALS2 + LTS, high percentile values (90%) for NIR and NDVI and interquartile range and standard deviation of SWIR and red were frequently selected. In the models combining ALS and LTS metrics, the LTS metrics appeared in the models with more than four predictors. Thus, LTS metrics have less predictive power for the GC, but once they are incorporated with ALS, they explain additional variability. Fig. 6 compares the cross-validated GC predictions obtained when using metrics from the single sensor alone (ALS1, ALS2, LTS) and for the combination of ALS and LTS, against those observed in the field plots. LTS

Table 2
Summary of the Gini coefficient modeling results based on different data sets and their combinations.

Data	Number of predictors	Prediction accuracy	Prediction precision		Agreement		Obs.-Pre. 1:1 correspondence		Overfitting
		rMD_{CV}	$RMSE_{CV}$	$rRMSE_{CV}$	R^2_{fit}	R^2_{CV}	alpha	beta	SSR
ALS1	4	0.01%	0.07	13.90%	0.55	0.49	0.03 ^{NS}	0.95 ^{NS}	1.07
ALS2	5	−0.02%	0.08	15.10%	0.48	0.40	0.05 ^{NS}	0.90 ^{NS}	1.08
LTS	6	−0.25%	0.09	16.50%	0.39	0.28	0.09 ^{NS}	0.84 ^{NS}	1.09
ALS1 + LTS	6	−0.14%	0.06	12.50%	0.66	0.59	0.03 ^{NS}	0.94 ^{NS}	1.10
ALS2 + LTS	6	0.06%	0.07	13.00%	0.62	0.55	0.03 ^{NS}	0.94 ^{NS}	1.09

rMD_{CV}: relative mean difference (Eq. (1)), $RMSE_{CV}$: leave-one-out cross-validation root mean square error (Eq. (2)), $rRMSE_{CV}$: relative $RMSE_{CV}$ (Eq. (3)), R^2_{fit}: coefficient of determination (Eq. (4)), R^2_{CV}: leave-one-out cross-validation R^2 (Eq. (5)), alpha and beta and their significance (Eq. (6)), SSR: sums of squares ratio (Eq. (7)). The best models for each data set are shown here (rest in the Appendix Table A.2). See explanation of metrics in Appendix Table A.1 and estimates of the model parameters in Appendix Table A.3. Levels of significance; *** < 0.001; ** < 0.01; * < 0.05; ˚ < 0.1; ^{NS} Non-significant.

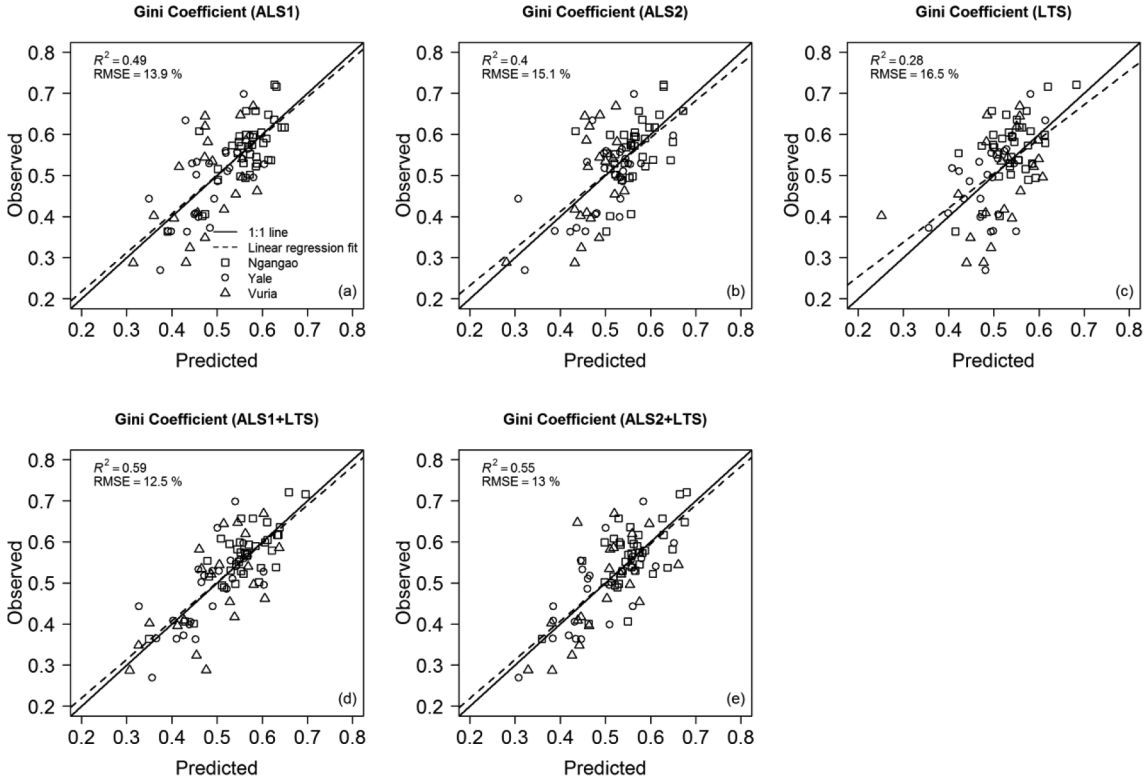

Fig. 6. Relationship between predicted (leave-one-out cross-validated) and observed Gini coefficient based on the best model using different metrics: (a) ALS1, (b) ALS2, (c) LTS, (d) ALS1 + LTS, and (e) ALS2 + LTS.

models tend to overestimate small values and underestimate large values (Fig. 6c).

3.3. Analysis of Gini coefficient maps

The best beta regression model (ALS1 + LTS) was applied to obtain spatially continuous predictions of GC for Ngangao, Yale, and Vuria (Fig. 7). The GC predictions revealed considerable differences among the three forest remnants and between the different forest types. Forest stands have even-sized, irregular, or bimodal diameter distribution of tree sizes within each forest remnant (Fig. 7). The exotic plantation forests of cypress and pine show low GC (typically < 0.5), meaning stands of even height, while indigenous and eucalyptus stands show high GC (typically > 0.5), meaning more uneven canopy. For example, a pine stand in the middle of Yale shows uniform canopy in Fig. 2b and a GC of less than 0.4 in Fig. 7, while the eucalyptus stands in Fig. 3b shows uneven canopy and a GC of more than 0.6 in Fig. 7. The uneven canopy structure of indigenous forests in Ngangao is shown in Fig. 2c, while in Fig. 7 it is represented as GC between 0.5 and 0.8. In addition, the uneven canopy structure of pine and cypress stands is seen, represented as GC between 0.1 and 0.5 in Fig. 7.

The distribution of GC in the indigenous, eucalyptus, cypress, and pine forests in each forest remnant is shown in Fig. 8. Yale showed relatively high values of GC, indicating emergent trees and several layers. On the other hand, the indigenous forest in Yale and Vuria showed GC prediction even below 0.30, indicating either a seedling stand or even-sized mature trees (Fig. 8b, c). Yale contained more eucalyptus forest than Ngangao or Vuria (Fig. 7). The eucalyptus forest showed higher values of GC than the other plantation forests (Table A4). This is seen in Fig. 2, in which the right hand top of (a) covered by eucalyptus shows the most uneven canopy.

Overall, tree size heterogeneity in the study area was above the $\widetilde{GC} > 0.5$ threshold for Ngangao and Yale (H_0: $\widetilde{GC} \leq 0.5$, H_1: $\widetilde{GC} > 0.5$, $p < 0.001$ for both), while it was less than 0.5 for Vuria (H_0: $\widetilde{GC} \geq 0.5$, H_1: $\widetilde{GC} < 0.5$, $p < 0.001$). The structural diversity in the indigenous and single-species plantation forest is diverse. Indigenous forests have bimodal (H_0: $\widetilde{GC} \leq 0.5$, H_1: $\widetilde{GC} > 0.5$, $p < 0.001$ for Ngangao) or even-sized distribution (H_0: $\widetilde{GC} \geq 0.5$, H_1:

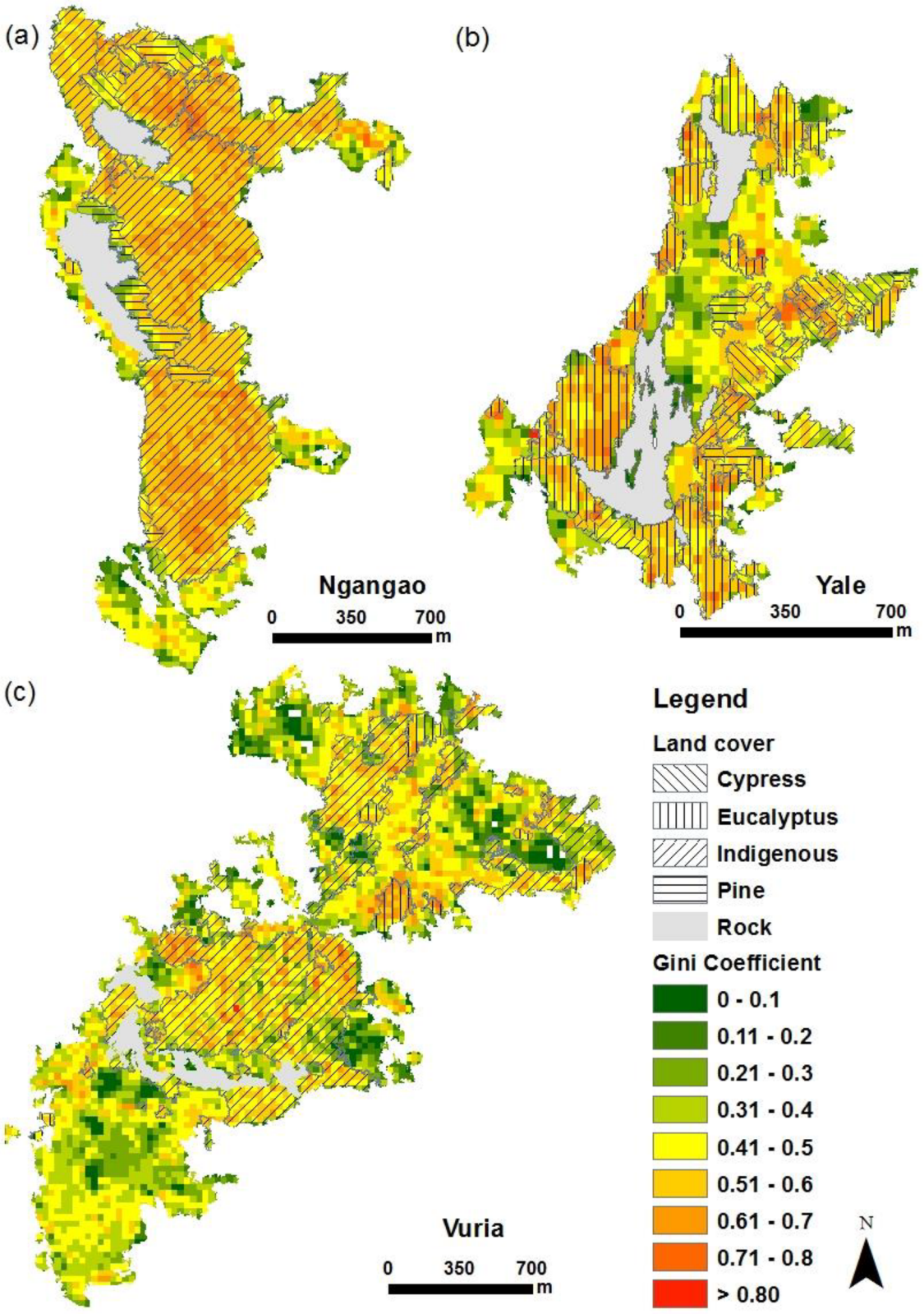

Fig. 7. Gini coefficient maps of (a) Ngangao, (b) Yale, and (c) Vuria. GC $<$ 0.5, GC = 0.5, and GC $>$ 0.5 represent even-sized forests, irregular forests, and multi-stratified forests, respectively. The map shows only the land cover types studied within the forest fragments.

$\widetilde{GC} < 0.5$, $p < 0.001$ for Vuria). However, even with the larger sample size provided by the GC maps, we failed to reject the null hypothesis for Yale (H_0: $\widetilde{GC} \geq 0.5$, H_1: $\widetilde{GC} < 0.5$, $p = 0.114$). The ranges of GC in the indigenous and eucalyptus forests (0.17–0.76 and 0.18–0.72) in Vuria were similar, representing higher structural diversity, however, their $\widetilde{GCs}$ were different. Due to the lower montane nature of the indigenous forest (Ind) in Ngangao, significantly higher $\widetilde{GC}$ was observed than in the eucalyptus forest (Euc) (H_0: $\widetilde{GC}_{Ind} \leq \widetilde{GC}_{Euc}$, H_1: $\widetilde{GC}_{Ind} > \widetilde{GC}_{Euc}$,

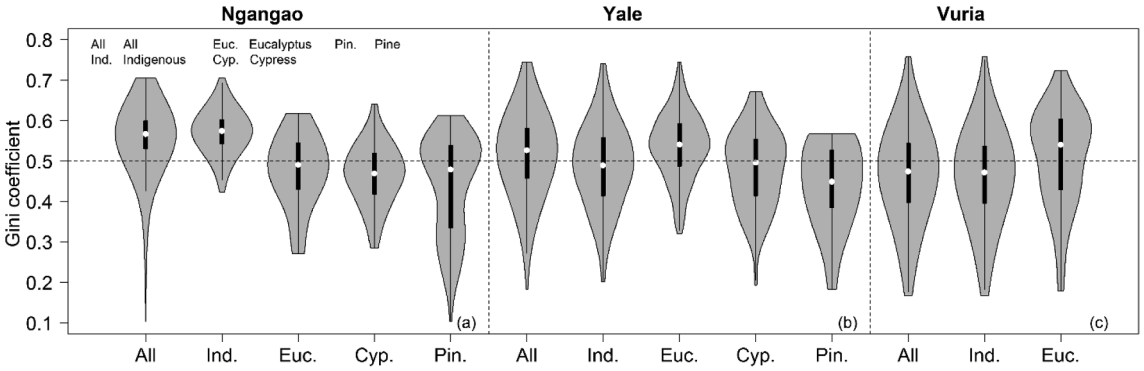

Fig. 8. Gini coefficient for indigenous and exotic plantation (eucalyptus, cypress, and pine) in the forest remnants in Taita Hills based on pixel values. The white dot represents the median, the thick black bar in the center represents the interquartile range, and the thin black line represents the rest of the distribution (1.5 × interquartile range). Horizontal line represents GC = 0.5.

$p < 0.001$), whereas in Yale and Vuria indigenous forests had significantly lower $\widetilde{GC}$ than the eucalyptus plantations (H_0: $\widetilde{GC}_{Ind} \geq \widetilde{GC}_{Euc}$, H_1: $\widetilde{GC}_{Ind} < \widetilde{GC}_{Euc}$, both $p < 0.001$).

In Yale, eucalyptus forests had significantly higher $\widetilde{GC}$ than cypress forests (*Cyp*), but not in Ngangao (H_0: $\widetilde{GC}_{Euc} \leqslant \widetilde{GC}_{Cyp}$, H_1: $\widetilde{GC}_{Euc} > \widetilde{GC}_{Cyp}$, $p < 0.001$ for Yale and $p = 0.406$ for Ngangao). A similar result was obtained when comparing eucalyptus and pine (H_0: $\widetilde{GC}_{Euc} \leqslant \widetilde{GC}_{Pine}$, H_1: $\widetilde{GC}_{Euc} > \widetilde{GC}_{Pine}$, $p < 0.001$ for Yale, and $p = 0.356$ for Ngangao). A similar outcome was obtained when evaluating whether cypress forests had significantly higher $\widetilde{GC}$ than pine forests (H_0: $\widetilde{GC}_{Cyp} \leqslant \widetilde{GC}_{Pine}$, H_1: $\widetilde{GC}_{Cyp} > \widetilde{GC}_{Pine}$, $p = 0.051$ for Yale, and $p = 0.332$ for Ngangao). The indigenous and eucalyptus forests in Ngangao, Yale, and Vuria were significantly different (H_0: $\widetilde{GC}_{Ind}^{Ngangao} = \widetilde{GC}_{Ind}^{Yale} = \widetilde{GC}_{Ind}^{Vuria}$, H_1: $\widetilde{GC}_{Ind}^{Ngangao} \neq \widetilde{GC}_{Ind}^{Yale} \neq \widetilde{GC}_{Ind}^{Vuria}$ and H_0: $\widetilde{GC}_{Euc}^{Ngangao} = \widetilde{GC}_{Euc}^{Yale} = \widetilde{GC}_{Euc}^{Vuria}$, H_1: $\widetilde{GC}_{Euc}^{Ngangao} \neq \widetilde{GC}_{Euc}^{Yale} \neq \widetilde{GC}_{Euc}^{Vuria}$, for both $p < 0.001$). This is evident since the indigenous forest is most intact in Ngangao, while in Vuria it is severely degraded due to selective logging, grazing in the forest, and planting of exotic trees within the indigenous stands.

4. Discussion

The results show that the GC of tree size heterogeneity can be predicted with reasonable accuracy, using ALS data and beta regression, in tropical montane forests. The resulting maps can be used to demonstrate the structural difference between forest remnants and forest types better than using field plots alone since they provide an idea of GC variation at the forest fragment level. The accuracy of the GC estimates improved significantly when using a combination of airborne LIDAR and satellite imagery (ALS + LTS), compared with predicting GC using either of these alone. Significant differences were found between forest stands of indigenous and plantation species within and between forest remnants, some of which could not be detected using the field information, which demonstrates the practical advantages of employing these data to study the ecology of the tree assemblages involved.

This was the first study carried out in the tropics showing how ALS can provide reliable and accurate predictions of GC. The accuracy of GC prediction depends on the pulse density (Adnan et al., 2017). ALS2 provided lower R^2_{CV} than ALS1, which can partially be explained by the lower point density as a result of the higher platform height. Furthermore, the lower frequency of returns may lead to a lower frequency of pulses detecting the understorey, underestimating the overall forest structure heterogeneity (Goodwin et al., 2006; Valbuena et al., 2017c). R^2 for GC prediction based on ALS alone in Valbuena et al., (2017b) was 0.21, which is much less than 0.40 for ALS2 and 0.49 for ALS1 in this study. The $rRMSE_{CV}$ obtained using ALS1 and ALS2 only (13.9% and 15.1%, respectively) is comparable with the corresponding figures reported in the recent studies by Valbuena et al. (2013a) (20.5%), Valbuena et al. (2013b) (19.7%), Valbuena et al. (2014) (16.8% and 16.1%), and Valbuena et al. (2016) (20.2% and 18.8%). Our results are similar to Goodwin et al. (2006), where higher platform altitudes underestimated forest structural properties. The difference in ALS sensor and platform height might have had an impact on the final model, warranting additional investigations. Wall-to-wall prediction of GC at grid sizes other than 30 m could have provided different conclusions but was consistent with 0.1 ha field plot size. Further research should, however, implement this method in predicting tree size inequality in other forest environments to test whether the ALS and Landsat combination yields similar synergies for other forest structure attributes.

The metrics derived from Landsat imagery, the LTS, alone showed little explanatory capacity for predicting forest heterogeneity. Valbuena et al. (2017b) predicted GC using a multispectral sensor (MS) alone with R^2 of 0.06, which is lower than 0.30 in our case. LTS is limited to spectral information from the top of the closed tree canopy and lacks vertical structural information. However, LTS has the potential to complement the information from the ALS in describing the structural complexity of the forest. Incorporation of spectral-temporal information from Landsat with ALS metrics improved prediction of tree size inequality. For example, with LTS metrics the inter-percentile range of NDVI and interquartile range of RSR showed significant explanatory potential for GC prediction. The key for these can be in the relationships between these remote sensing metrics and the forest density (Adhikari et al., 2016) since Adnan et al. (2017) showed that stand density is a confounding factor in the relationships between ALS metrics and the GC of tree size heterogeneity. This is possibly the reason why Manzanera et al. (2016) detected good potential in NDVI metrics for predicting GC. Our study is the first to exploit such potential with satellite imagery of coarser resolution. Also, by including LTS metrics sensitive to tree species with ALS2 metrics, GC prediction was improved relative to prediction without LTS. This shows that even with lower investments on ALS measurements and including LTS metrics, we can achieve higher R^2_{CV}. Valbuena et al. (2017b) achieved R^2_{CV} of 0.45 when combining metrics from ALS and MS (NDVI), which is lower than what we achieved (0.59) by including inter-percentile range NDVI and interquartile range RSR along with ALS metrics (Model 18 in Table A2).

Our results suggest that if heterogeneity of tree sizes must be determined using ALS alone, a high point density is needed. In other cases, low-density ALS data sets could be supported by satellite data providing information about variation in forest density to obtain reliable results. If LTS metrics are included along with ALS sensor metrics, an ALS campaign with low point density may suffice to get higher R^2_{CV} and lower $rRMSE_{CV}$. LTS data can thus compensate for lower point density. Combining ALS1 with LTS reduced $rRMSE_{CV}$ to 12.5%. Synergistic use of ALS and LTS has been documented for mapping biomass (Phua et al., 2017), canopy height (Hudak et al., 2002), and canopy cover (Ahmed et al., 2014), but this is the first study to map forest structure

heterogeneity based on GC using ALS and LTS in the tropics.

ALS and LTS metrics-based wall-to-wall prediction of GC provides an opportunity to visualize tree size inequality among different forest types and stands. Based on a range of GC predication, different forest stands within Ngangao, Yale, and Vuria have even-sized, irregular, and bimodal diameter distributions for indigenous and exotic plantation species. As a result, we were able to make comparisons and analyse the ecology of the species in these forests that we would have been unable to do from field information alone.

During 1955 and 2004, a total of 260 ha (50%) of indigenous tropical forest were deforested and degraded by conversion into cropland and bushland. In the same period, plantation and indigenous forest were established on the barren lands, resulting in a balanced total forest area (Pellikka et al., 2009). In 1955 indigenous forest developed in non-forested areas, and some areas were covered by the fast growth of succession species, e.g. *Phoenix reclinata* in abandoned open areas (Pellikka et al., 2009). Furthermore, indigenous species, e.g. *Ocotea usambarensis* and *Podocarpus latifolius,* were extracted from Ngangao for timber in the past, leaving small diameter classes (Aerts et al., 2011). For these reasons, the indigenous forest is a mixture of climax forest (undisturbed) with bimodal diameter distribution and primary successions with even-sized diameter distributions.

Our results were similar to those obtained by Valbuena et al. (2016), where old protected areas similar to the indigenous forest had the highest GC values. Eucalyptus, cypress, and pine were planted between the 1950 s and 1970 s in non-forested areas or degraded areas due to deforestation. As now mature stands were planted, it is evident that their canopy structure is of even height relative to indigenous forests, whose structures were more governed by disturbance ecology (Mbuthia, 2003). These are fast-growing species and can be an alternative to reforestation projects.

It is important to emphasize that many of these conclusions that we were able to draw from the remote sensing prediction map of GC could not be concluded using the field plot information alone. For example, at the level of plot data, the indigenous forest has bimodal diameter distribution in all forest remnants. However, when using the remote sensing prediction, we were able to prove that the indigenous forest in Ngangao is bimodal and in Vuria even-sized. These led us to the inference about their distinctive ecology and historical land use, details that would not have come to light without the assistance of remote sensing. Moreover, the exotic forest (eucalyptus) in Vuria has bimodal diameter distribution at the remote sensing level, while at plot level the exotic forest is even-sized. Furthermore, the forest of Yale has a bimodal distribution, while Vuria is even-sized, which was not possible to statistically prove based on field plot information only due to a much smaller sample size that can be employed from discrete plot samples. We thus emphasize the potential of RS for increasing the statistical power in pursuing ecological hypotheses and showing patterns on landscape scales that cannot be detected otherwise.

5. Conclusion

We explored the reliability of predicting the GC of tree size heterogeneity in indigenous tropical montane forests and compared the results with those obtained in plantations of eucalyptus, cypress, and pine in the same area. For this purpose, we employed two ALS data sets acquired from different altitudes using two different sensors, and also explored the potential for improving the accuracy of the GC models by incorporating a set of LTS predictors based on Landsat time series. The results showed the inclusion of LTS spectral-temporal metrics sensitive to tree species and forest density improved the modeling accuracy significantly. The inclusion of NDVI inter-percentile range and RSR interquartile range along with ALS metrics had the highest R^2_{CV} among the models. Finally, we predicted maps of GC for the three forest remnants using the best model and compared forest structural heterogeneity among and within the forest types. Results showed that these maps can be employed to demonstrate ecological hypotheses that would not be plausible using field data alone. Using these GC maps, we observed a high degree of structural heterogeneity in these montane forest remnants. Based on the ranges of GC values observed in these maps, we were able to identify different areas within these forest remnants having even-sized, irregular, and bimodal diameter distributions for indigenous and exotic species. For modeling GC, high point density ALS is required if ALS metrics is used alone. Nonetheless, low scanning densities may suffice when combining the ALS data with LTS metrics. The method showed the potential to discriminate between indigenous and plantation forests.

Acknowledgements

We acknowledge the contributions of the BIODEV project funded by the Ministry for Foreign Affairs of Finland and the TAITAWATER, TAITASMART, and ECHOES projects funded by the Academy of Finland. The first author received funding from the Doctoral Programme in Atmospheric Sciences (ATM-DP) of the University of Helsinki. The research was approved by the National Council for Science and Technology, Kenya (permit no. NCST/RCD/17/012/33). The Taita Research Station of the University of Helsinki is gratefully acknowledged for logistical support, and Darius Kimuzi, Benson Mwakachola Lombo, Mwadime Mjomba, Jesse Hietanen, Elisa Schäfer, Vuokko Heikinheimo, and Jessica Broas as well as geography students of the Taita Field Course of the University of Helsinki in 2018 are thanked for field assistance.

Appendix

Modeling results

Table A.1 explains the predictor metrics used in modeling GC. Fig. A.1 shows R^2_{CV} for GC prediction using a different number of variables. The highest R^2_{CV} obtained using six metrics together from ALS1 and LTS metrics was 0.59. Maximum predictors in ALS1 were four and in ALS2 five; the R^2_{CV} for four metrics in ALS1 are still higher than five metrics in ALS2 (Fig. A.1, Table A.2). Table A.2 shows all of the models with different numbers of metrics possible in each sensor combination or alone. The estimates of the model coefficients (beta distribution parameters, β and$\varnothing$) are given in Table A.3. In the final choice of models, we selected those with lower uncertainty while securing their significance (both in model fit – Table A.2 – and observed versus predicted – Eq. (6)) and avoiding overfitting to the sample (SSR, Eq. (7)). To avoid overfitting of the models, models with SSR greater than 1.1 were removed. Table A.4 presents the summary statistics of different forest pixels (indigenous, eucalyptus, cypress, and pine) in the three forest remnants.

Appendix: analysis of Gini coefficient maps

Table A.1
Summary of airborne laser scanning (ALS) metrics computed using Fusion (McGaughey 2016) and spectral-temporal metrics from the Landsat time series (LTS).

Data set	Predictor	Description
ALS[1]	H.p01, H.p05, H.p10, H.p20, H.p25, H.p30, H.p40, H.p50, H.p60, H.p70, H.p75, H.p80, H.p90, H.p95, H.p99	1st, 5th, 10th … and 99th percentile of return height > 3 m
	H.L1, H.L2, H.L3, H.L4	L-moments 1–4 of return height > 3 m
	H.L.cv	L-moments coefficient of variation of return height > 3 m
	H.L.skewness	L-moments skewness of return height > 3 m
	H.L.kurtosis	L-moments kurtosis of return height > 3 m
	H.max	Maximum of return height > 3 m
	H.mean	Mean of return height > 3 m
	H.mode	Mode of return height > 3 m
	H.cv	Coefficient of variation of return height > 3 m
	H.v	Variance of return height > 3 m
	H.stdev	Standard deviation of return height > 3 m
	H.skewness	Skewness of return height > 3 m
	H.kurtosis	Kurtosis of return height > 3 m
	H.IQ	75th percentile minus 25th percentile for cell
	CC.first	First returns > 3 m/Total first returns * 100
	CC.all	All returns > 3 m/Total all returns * 100
	CC.all.first	All returns > 3 m/Total first returns * 100
	CC.first.mean	First returns above mean/Total first returns * 100
	CC.all.mean	All returns above mean/Total all returns * 100
	CC.all.mean.first	All returns above mean/Total first returns * 100
	CC.first.mode	First returns above mode/Total first returns * 100
	CC.all.mode	All returns above mode/Total all returns * 100
	CC.all.mode.first	All returns above mode/Total first returns * 100
LTS[2]	B2…B7 (Blue, Green, Red, NIR, SWIR1, SWIR2)	Reflectance in Landsat 8 OLI blue, green, red, near infrared and two shortwave infrared spectral bands
	NDVI	(NIR − Red)/(NIR + Red) (Rouse et al., 1973)
	RSR	(NIR/Red) × [(SWIR1_max − SWIR1)/(SWIR1_max − SWIR1_min)] (Brown et al., 2000)
	B, G, W	Tasseled cap brightness, greenness, and wetness (Crist 1985)

[1] For ALS, all height metrics (start with 'H') were calculated separately using first and last pulse returns. All canopy cover metrics (start with "CC") were calculated using all (single, first and last) returns only.

[2] For LTS different percentiles, trimmed mean and inter-percentile range were calculated. Prefix "B2_", "B3_", "B4_", "B5_", "B6_", and " B7_" represents Blue, Green, Red, NIR, SWIR1, and SWIR2 reflectance bands in Landsat 8 OLI. Prefixes "NDVI_", "RSR_", "B_", "G_", and "W_" represent normalized difference vegetation index (NDVI) (Rouse et al. 1973), reduced simple ratio (RSR) (Brown et al. 2000), and Tasseled cap brightness, greenness, and wetness (Crist 1985), respectively.

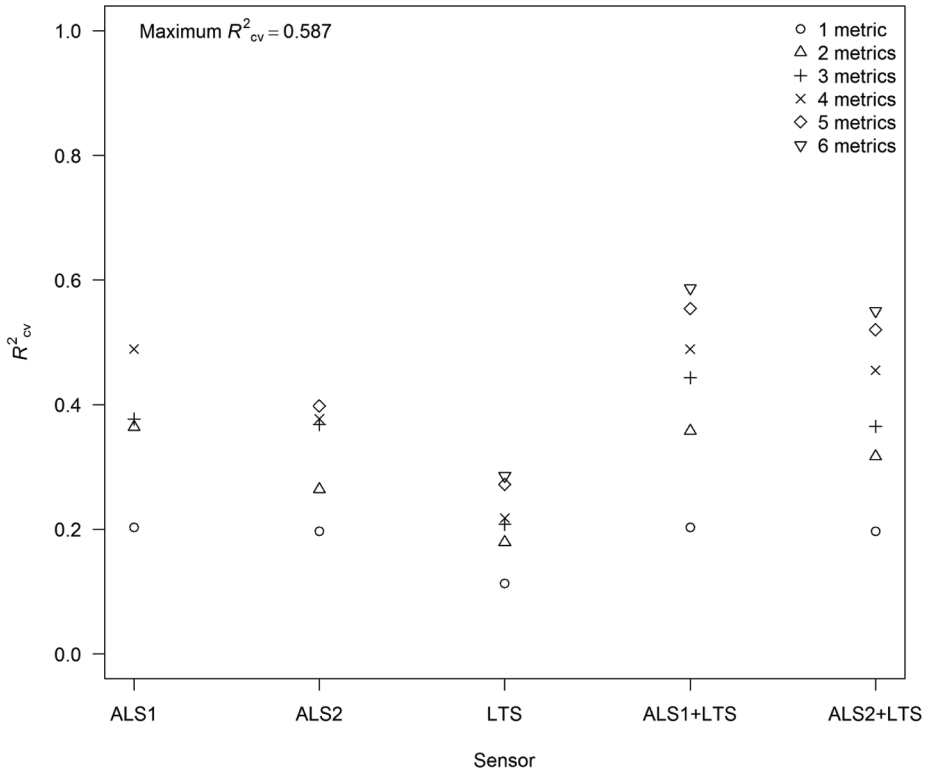

Fig. A.1. Cross-validated coefficient of determination (R^2_{CV}) for beta regression models with a range of metrics (maximum of six metrics).

Table A.2
Summary of the Gini coefficient modelling results based on different data sets and their combinations.

Model number	Data	Predictors	Model (Gini Coefficient ~)	Prediction precision			Agreement		Hypothesis test		Overfitting
				rMD_{CV}	$RMSE_{CV}$	$rRMSE_{CV}$	R^2_{fit}	R^2_{CV}	alpha	beta	SSR
1	ALS1	1	$b_0 + b_1 \times$ CC.all	−0.14	0.09	17.30	0.24	0.20	0.04 ^{NS}	0.93^{NS}	1.02
2	ALS1	2	$b_0 + b_1 \times$ CC.all + $b_2 \times$ LR_H.cv	0.02	0.08	15.50	0.41	0.36	0.03 ^{NS}	0.94 ^{NS}	1.04
3	ALS1	3	$b_0 + b_1 \times$ CC.all + $b_2 \times$ LR_H.cv + $b_3 \times$ LR_H.L4	0.02	0.08	15.30	0.44	0.37	0.04 ^{NS}	0.92 ^{NS}	1.06
4	**ALS1**	**4**	$\boldsymbol{b_0 + b_1 \times}$ **CC.all** $\boldsymbol{+ b_2 \times}$ **FR_H.cv** $\boldsymbol{+ b_3 \times}$ **FR_H.p05** $\boldsymbol{+ b_4 \times}$ **LR_H.p05**	**0.01**	**0.07**	**13.90**	**0.55**	**0.49**	**0.03** ^{NS}	**0.95** ^{NS}	**1.07**
5	ALS2	1	$b_0 + b_1 \times$ CC.all.first	−0.13	0.09	17.40	0.24	0.19	0.06 ^{NS}	0.89 ^{NS}	1.03
6	ALS2	2	$b_0 + b_1 \times$ CC.all + $b_2 \times$ LR_H.cv	−0.05	0.09	16.70	0.34	0.26	0.06 ^{NS}	0.88 ^{NS}	1.06
7	ALS2	3	$b_0 + b_1 \times$ FR_H.L.kurtosis + $b_2 \times$ FR_H.p01 + $b_3 \times$ LR_H.kurtosis	−0.10	0.08	15.70	0.43	0.37	0.04 ^{NS}	0.93 ^{NS}	1.05
8	ALS2	4	$b_0 + b_1 \times$ FR_H.L.skewness + $b_2 \times$ FR_H.L.kurtosis + $b_3 \times$ FR_H.p01 + $b_4 \times$ LR_H.kurtosis	−0.10	0.08	15.60	0.45	0.37	0.05 ^{NS}	0.9 ^{NS}	1.07
9	**ALS2**	**5**	$\boldsymbol{b_0 + b_1 \times}$ **FR_H.L3** $\boldsymbol{+ b_2 \times}$ **FR_H.L.kurtosis** $\boldsymbol{+ b_3 \times}$ **FR_H.p01** $\boldsymbol{+ b_4 \times}$ **FR_H.p99** $\boldsymbol{+ b_5 \times}$ **LR_H.kurtosis**	**−0.02**	**0.08**	**15.10**	**0.48**	**0.40**	**0.05** ^{NS}	**0.9** ^{NS}	**1.08**
10	LTS	1	$b_0 + b_1 \times$ NDVI_p90$_t$	−0.21	0.10	18.30	0.15	0.11	0.08 ^{NS}	0.86 ^{NS}	1.02
11	LTS	2	$b_0 + b_1 \times$ B5_range10_90$_n$ + $b_2 \times$ NDVI_p10$_t$	−0.19	0.09	17.60	0.23	0.17	0.07 ^{NS}	0.87 ^{NS}	1.03
12	LTS	3	$b_0 + b_1 \times$ NDVI_p90$_t$ + $b_2 \times$ sd_B6_range25_75$_n$ + $b_3 \times$ sd_NDVI_p10$_t$	−0.22	0.09	17.30	0.26	0.21	0.05 ^{NS}	0.9 ^{NS}	1.03
13	LTS	4	$b_0 + b_1 \times$ B3_range10_90$_n$ + $b_2 \times$ B6_range10_90$_n$ + $b_3 \times$ NDVI_range10_90$_t$ + $b_4 \times$ RSR_range25_75$_t$	−0.21	0.09	17.20	0.32	0.21	0.10 ^{NS}	0.82 ^{NS}	1.08
14	LTS	5	$b_0 + b_1 \times$ B4_range25_75$_n$ + $b_2 \times$ RSR_range25_75$_t$ + $b3 \times$ Bri_p10$_n$ + $b_4 \times$ sd_B6_range25_75$_n$ + $b_5 \times$ sd_NDVI_p10$_t$	−0.24	0.09	16.60	0.36	0.26	0.08 ^{NS}	0.86 ^{NS}	1.08
15	**LTS**	**6**	$\boldsymbol{b_0 + b_1 \times}$ **B2_p90$_n$** $\boldsymbol{+ b_2 \times}$ **B4_range25_75$_n$** $\boldsymbol{+ b_3 \times}$ **B6_range10_90$_n$** $\boldsymbol{+ b_4 \times}$ **RSR_range25_75$_t$** $\boldsymbol{+ b_5 \times}$**sd_B6_range25_75$_n$** $\boldsymbol{+ b_6 \times}$**sd_NDVI_range10_90$_t$**	**−0.25**	**0.09**	**16.50**	**0.39**	**0.28**	**0.09** ^{NS}	**0.84** ^{NS}	**1.09**
16	ALS1 + LTS	3	$b_0 + b_1 \times$ CC.all + $b_2 \times$ LR_H.L.cv + $b_3 \times$ B5_p90$_n$	−0.02	0.08	14.50	0.50	0.44	0.03 ^{NS}	0.94 ^{NS}	1.06
17	ALS1 + LTS	5	$b_0 + b_1 \times$ CC.all + $b_2 \times$ FR_H.cv + $b_3 \times$ FR_H.p05 + $b_4 \times$ LR_H.p05 + $b_5 \times$ B5_p90$_t$	−0.04	0.07	13.00	0.62	0.55	0.03 ^{NS}	0.94 ^{NS}	1.09
18	**ALS1 + LTS**	**6**	$\boldsymbol{b_0 + b_1 \times}$ **CC.all** $\boldsymbol{+ b_2 \times}$ **FR_H.cv** $\boldsymbol{+ b_3 \times}$ **FR_H.p05** $\boldsymbol{+ b_4 \times}$ **LR_H.p05** $\boldsymbol{+ b_5 \times}$ **NDVI_range10_90$_t$** $\boldsymbol{+ b_6 \times}$**RSR_range25_75$_t$**	**−0.14**	**0.06**	**12.50**	**0.66**	**0.59**	**0.03** ^{NS}	**0.94** ^{NS}	**1.10**
19	ALS2 + LTS	2	$b_0 + b_1 \times$ CC.all.first + $b_2 \times$ B5_p90$_n$	−0.13	0.08	16.00	0.37	0.32	0.04 ^{NS}	0.92 ^{NS}	1.04
20	ALS2 + LTS	3	$b_0 + b_1 \times$ CC.all + $b_2 \times$ LR_H.L.CV + $b_3 \times$ B5_p90$_n$	−0.10	0.08	15.50	0.44	0.36	0.05 ^{NS}	0.91 ^{NS}	1.07
21	ALS2 + LTS	4	$b_0 + b_1 \times$ FR_H.L.kurtosis + $b_2 \times$ FR_H.p01 + $b_3 \times$ LR_H.kurtosis + $b_4 \times$ B5_p90$_n$	−0.04	0.08	14.60	0.52	0.45	0.04 ^{NS}	0.93 ^{NS}	1.07
22	ALS2 + LTS	5	$b_0 + b_1 \times$ FR_H.L.kurtosis + $b_2 \times$ FR_H.p01 + $b_3 \times$ LR_H.kurtosis + $b_4 \times$ NDVI_p90$_t$ + $b_5 \times$ sd_B6_ range25_75$_n$	−0.07	0.07	13.70	0.58	0.52	0.03 ^{NS}	0.94 ^{NS}	1.07
23	**ALS2 + LTS**	**6**	$\boldsymbol{b_0 + b_1 \times}$ **FR_H.L.kurtosis** $\boldsymbol{+ b_2 \times}$ **FR_H.p01** $\boldsymbol{+ b_3 \times}$ **LR_H.kurtosis** $\boldsymbol{+ b_4 \times}$ **LR_H.L.CV** $\boldsymbol{+ b_5 \times}$ **NDVI_p90$_t$** $\boldsymbol{+ b_6 \times}$ **sd_B4_ range25_75$_n$**	**0.06**	**0.07**	**13.00**	**0.62**	**0.55**	**0.03** ^{NS}	**0.94** ^{NS}	**1.09**

rMD_{CV} :relative mean difference (Eq. (1)), $RMSE_{CV}$: leave–one–out–cross–validation root-mean-square-error (Eq. (2)), $rRMSE_{CV}$: relative $RMSE_{CV}$ (Eq. (3)), R^2_{fit}: Pseudo coefficient of determination (Eq. (4)), R^2_{CV}(leave–one–out–cross–validation R^2 (Eq. (5)), alpha and beta and their significance (Eq. (6)), SSR: sums of squares ratio (Eq. (7)). The best models for each data set are shown in bold. Subscript "t" refers to the index or band without and subscript "n" with topographic normalization based on JAXA DEM. All height metrics (start with 'H') were calculated separately using first and last pulse returns, which is indicated by a prefix, 'FR_' or 'LR_', respectively. Landsat Time Series included 10th, 25th, 50th, 75th, and 90th percentiles, trimmed mean (10% and 25%), inter-percentile range (10–90), and interquartile (25–75), which is indicated by a suffix, 'p10', 'p25', 'p50', 'p75', 'p90', 'tm10', 'tm25', 'range10_90', and 'range25_75', respectively. Prefix "sd_" represents standard deviation of all of these metrics. See explanation of metrics in Appendix Table A.1 and estimates of the model parameters in Appendix Table A.3. Levels of significance; *** < 0.001; ** < 0.01; * < 0.05; $^{\circ}$ < 0.1; ^{NS} Non-significant.

Table A.3
Estimates of model parameters (Table A.2). β_0, β_1, β_2, β_3, β_4, β_5, β_6, and ∅ are model-specific coefficients for model numbers 1 to 23.

Model number	Sensor data	Estimates for parameters[a]							
		β_0	β_1	β_2	β_3	β_4	β_5	β_6	∅
1	ALS1	−0.586**	0.012***						31.17***
2	ALS1	−1.4841***	0.0175***	1.229***					40.43***
3	ALS1	−1.6462***	0.0178***	1.4461***	0.241*				42.61***
4	**ALS1**	**−1.6345*** **	**0.0184*** **	**1.6689*** **	**0.0361*** **	**−0.0456*** **			**52.99*** **
5	ALS2	−0.565**	0.008***						31.35***
6	ALS2	−2.15***	0.0234***	1.3906***					35.79***
7	ALS2	0.2272***	1.3396***	0.0489***	−0.0826***				41.7***
8	ALS2	0.2261***	0.3229°	1.5861***	0.0515***	−0.0875***			43.53***
9	**ALS2**	**0.0597°**	**0.1453***	**1.4601*** **	**0.0418*** **	**0.0078****	**−0.0813*** **		**45.96*** **
10	LTS	−2.3327**	3.2663***						27.97***
11	LTS	−1.193***	0.0004***	1.7761***					30.7***
12	LTS	−2.1284**	2.9387**	0.0044***	−4.3161*				31.9***
13	LTS	0.3725***	−0.0056***	0.0015***	−3.5277***	0.1757***			34.68***
14	LTS	−0.0723°	−0.0074***	0.1442***	0.0002**	0.0055***	−4.4606**		37.28***
15	**LTS**	1.3141***	−0.0053***	−0.0047*	0.0008*	0.1588***	0.0042***	−5.6755**	38.94***
16	ALS1 + LTS	−1.9671***	0.0165***	2.1529***	0.0002***				47.25***
17	ALS1 + LTS	−2.0564***	0.0175***	1.6331***	0.0361***	−0.0447***	0.0002***		62.53***
18	**ALS1 + LTS**	**−1.7158*** **	**0.0153*** **	**2.3563*** **	**0.0404*** **	**−0.0402*** **	**−1.9315*** **	**0.1203*** **	**68.85*** **
19	ALS2 + LTS	−1.2636***	0.0083***	0.0002***					37.5***
20	ALS2 + LTS	−2.6384***	0.0221***	2.4585***	0.0002***				42.2***
21	ALS2 + LTS	−0.3636*	1.4187***	0.0465***	−0.0808***	0.0002***			49.73***
22	ALS2 + LTS	−2.1133***	1.4924***	0.0455***	−0.0818***	2.5842***	0.0033***		56.73***
23	**ALS2 + LTS**	**−2.6696*** **	**1.7608*** **	**0.0624*** **	**−0.0705*** **	**1.7422****	**2.6733*** **	**0.0073*** **	**62.26*** **

Levels of significance; *** < 0.001; ** < 0.01; * < 0.05; ° < 0.1; [NS] Non-significant. Independent metrics which correspond to each beta coefficient (β) are detailed in Table A.2

Table A.4
Summary statistics for the three forests and forest types based on the pixels in the maps. $\overline{GC}$ is mean, σ is standard deviation, and IQR is interquartile range of GC.

Land cover	Ngangao					Yale					Vuria				
	$\overline{GC}$	range	σ (GC)	IQR	median	$\overline{GC}$	range	σ (GC)	IQR	median	$\overline{GC}$	range	σ (GC)	IQR	median
Indigenous	0.57	0.42–0.71	0.05	0.06	0.57	0.49	0.20–0.74	0.10	0.14	0.49	0.46	0.17–0.76	0.11	0.14	0.47
Eucalyptus	0.47	0.27–0.62	0.09	0.12	0.49	0.54	0.32–0.74	0.08	0.11	0.54	0.51	0.18–0.72	0.12	0.18	0.54
Cypress	0.47	0.28–0.64	0.08	0.10	0.47	0.48	0.19–0.67	0.10	0.14	0.50					
Pine	0.44	0.10–0.61	0.12	0.20	0.48	0.44	0.18–0.57	0.10	0.14	0.45					
All	0.57	0.10–0.71	0.05	0.06	0.57	0.49	0.18–0.74	0.10	0.14	0.49	0.46	0.17–0.76	0.11	0.14	0.47

References

Adhikari, H., Heiskanen, J., Maeda, E.E., Pellikka, P.K.E., 2015. Does topographic normalization of landsat images improve fractional tree cover mapping in tropical mountains? Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XL-7/W3, 261–267. https://doi.org/10.5194/isprsarchives-XL-7-W3-261-2015.

Adhikari, H., Heiskanen, J., Maeda, E.E., Pellikka, P.K.E., 2016. The effect of topographic normalization on fractional tree cover mapping in tropical mountains: an assessment based on seasonal Landsat time series. Int. J. Appl. Earth Obs. Geoinf. 52, 20–31. https://doi.org/10.1016/j.jag.2016.05.008.

Adhikari, H., Heiskanen, J., Siljander, M., Maeda, E., Heikinheimo, V., Pellikka, K.E., P., 2017. Determinants of aboveground biomass across an Afromontane Landscape Mosaic in Kenya. Remote Sens. 9 (8), 827. https://doi.org/10.3390/rs9080827.

Adnan, S., Maltamo, M., Coomes, D.A., Valbuena, R., 2017. Effects of plot size, stand density, and scan density on the relationship between airborne laser scanning metrics and the Gini coefficient of tree size inequality. Can. J. For. Res. 47 (12), 1590–1602. https://doi.org/10.1139/cjfr-2017-0084.

Aerts, R., Thijs, K.W., Lehouck, V., Beentje, H., Bytebier, B., Matthysen, E., Gulinck, H., Lens, L., Muys, B., 2011. Woody plant communities of isolated Afromontane cloud forests in Taita Hills, Kenya. Plant Ecol. 212 (4), 639–649. https://doi.org/10.1007/s11258-010-9853-3.

Ahmed, O.S., Franklin, S.E., Wulder, M.A., 2014. Integration of LiDAR and Landsat data to estimate forest canopy cover in Coastal British Columbia. Photogramm. Eng. Remote Sens. 80 (10), 953–961. https://doi.org/10.14358/PERS.80.10.953.

Almeida, Stark, S.C., Schietti, J., Camargo, J.L.C., Amazonas, N.T., Görgens, E.B., Rosa, D. M., Smith, M.N., Valbuena, R, Saleska, S., Nelson, B.W., Andrade, A., Mesquita R., Laurance, S.G., Laurance, W.F., T.E., L., & P.H.S, B., 2019. Persistent Effects of Fragmentation on Tropical Rainforest Canopy Structure after 20 Years of Isolation. Ecological Applications., [in press].

Bergen, K.M., Goetz, S.J., Dubayah, R.O., Henebry, G.M., Hunsaker, C.T., Imhoff, M.L., Nelson, R.F., Parker, G.G., Radeloff, V.C., 2009. Remote sensing of vegetation 3-D structure for biodiversity and habitat: review and implications for lidar and radar spaceborne missions. J. Geophys. Res.: Biogeosci. 114 (G2). https://doi.org/10.1029/2008JG000883.

Boström, M., 2010. Ngangao, Taita Hills, Kenya, Scale 1:10000. University of Helsinki, Finland (Unpublished).

Boström, M., 2013. Tree stand map Vuria, Taita Hills, Kenya, Scale 1:10000. University of Helsinki, Finland (Unpublished).

Brown, L., Chen, J.M., Leblanc, S.G., Cihlar, J., 2000. A shortwave infrared modification to the simple ratio for LAI retrieval in boreal forests: an image and model analysis. Remote Sens. Environ. 71 (1), 16–25. https://doi.org/10.1016/S0034-4257(99)00035-8.

Burgess, N.D., Butynski, T.M., Cordeiro, N.J., Doggart, N.H., Fjeldså, J., Howell, K.M., Kilahama, F.B., Loader, S.P., Lovett, J.C., Mbilinyi, B., Menegon, M., Moyer, D.C., Nashanda, E., Perkin, A., Rovero, F., Stanley, W.T., Stuart, S.N., 2007. The biological importance of the Eastern Arc Mountains of Tanzania and Kenya. Biol. Conserv. 134 (2), 209–231. https://doi.org/10.1016/j.biocon.2006.08.015.

Camargo, J.L.C., Kapos, V., 2009. Complex edge effects on soil moisture and microclimate in central Amazonian forest. J. Trop. Ecol. 11 (2), 205–221. https://doi.org/10.1017/S026646740000866X.

Clawges, R., Vierling, K., Vierling, L., Rowell, E., 2008. The use of airborne lidar to assess avian species diversity, density, and occurrence in a pine/aspen forest. Remote Sens. Environ. 112 (5), 2064–2073. https://doi.org/10.1016/j.rse.2007.08.023.

Coops, N.C., Tompaski, P., Nijland, W., Rickbeil, G.J.M., Nielsen, S.E., Bater, C.W., Stadt, J.J., 2016. A forest structure habitat index based on airborne laser scanning data. Ecol. Ind. 67, 346–357. https://doi.org/10.1016/j.ecolind.2016.02.057.

Cribari-Neto, F., Zeileis, A., 2010. Beta Regression in R. Journal of Statistical Software, 34(2), 1-24. http://www.jstatsoft.org/v34/i02/.

Crist, E.P., 1985. A TM Tasseled Cap equivalent transformation for reflectance factor data. Remote Sens. Environ. 17 (3), 301–306. https://doi.org/10.1016/0034-4257(85)90102-6.

Dalponte, M., Ene, L.T., Gobakken, T., Næsset, E., & Gianelle, D., 2018. Predicting Selected Forest Stand Characteristics with Multispectral ALS Data, 10(4), 586. DOI:

10.3390/rs10040586.
Erdelen, M., 1984. Bird Communities and Vegetation Structure: I. Correlations and Comparisons of Simple and Diversity Indices. Oecologia 61 (2), 277–284.
Erdogan, H.E., Pellikka, P.K.E., Clark, B., 2011. Modelling the impact of land-cover change on potential soil loss in the Taita Hills, Kenya, between 1987 and 2003 using remote-sensing and geospatial data. Int. J. Remote Sens. 32 (21), 5919–5945. https://doi.org/10.1080/01431161.2010.499379.
Erfanifard, Y., Stereńczak, K., Miścicki, S., 2019. Management strategies alter competitive interactions and structural properties of Norway spruce in mixed stands of Bialowieża Forest, Poland. For. Ecol. Manage. 437, 87–98. https://doi.org/10.1016/j.foreco.2019.01.035.
ESRI, 2015. Environmental Systems Research Institute Inc.; Redlands, CA: 1999–2015. ArcGIS Basemap Data.
Goetz, S., Steinberg, D., Dubayah, R., Blair, B., 2007. Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA. Remote Sens. Environ. 108 (3), 254–263. https://doi.org/10.1016/j.rse.2006.11.016.
Goodwin, N.R., Coops, N.C., Culvenor, D.S., 2006. Assessment of forest structure with airborne LiDAR and the effects of platform altitude. Remote Sens. Environ. 103 (2), 140–152. https://doi.org/10.1016/j.rse.2006.03.003.
Gorgens, E.B., Valbuena, R., Rodriguez, L.C.E., 2017. A method for optimizing height threshold when computing airborne laser scanning metrics. Photogramm. Eng. Remote Sens. 83 (5), 343–350. https://doi.org/10.14358/PERS.83.5.343.
Hansen, M.C., Potapov, P.V., Goetz, S.J., Turubanova, S., Tyukavina, A., Krylov, A., Kommareddy, A., Egorov, A., 2016. Mapping tree height distributions in Sub-Saharan Africa using Landsat 7 and 8 data. Remote Sens. Environ. 185, 221–232. https://doi.org/10.1016/j.rse.2016.02.023.
Heiskanen, J., Korhonen, L., Hietanen, J., Pellikka, P.K.E., 2015. Use of airborne lidar for estimating canopy gap fraction and leaf area index of tropical montane forests. Int. J. Remote Sens. 36 (10), 2569–2583. https://doi.org/10.1080/01431161.2015.1041177.
Heiskanen, J., Adhikari, H., Piiroinen, R., Packalen, P., Pellikka, P.K.E., 2019. Do airborne laser scanning biomass prediction models benefit from Landsat time series, hyperspectral data or forest classification in tropical mosaic landscapes? Int. J. Appl. Earth Obs. Geoinf. 81, 176–185. https://doi.org/10.1016/j.jag.2019.05.017.
Hill, R.A., Shelley, A., Hinsley, A., Broughton, R.K., 2014. Assessing habitats and organism-habitat relationships by airborne laser scanning. In: Maltamo, M., Næsset, E., Vauhkonen, J. (Eds.), Forestry Applications of Airborne Laser Scanning. Concepts and Case Studies. Springer, Dordrecht, pp. 269–292 Managing Forest Ecosystems Series 27.
Hudak, A.T., Lefsky, M.A., Cohen, W.B., Berterretche, M., 2002. Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height. Remote Sens. Environ. 82 (2–3), 397–416. https://doi.org/10.1016/S0034-4257(02)00056-1.
Isenburg, M., 2014. "LAStools - efficient LiDAR processing software" (version 141017, unlicensed), obtained from http://rapidlasso.com/LAStools.
JAXA, 2015. http://global.jaxa.jp/press/2015/05/20150518_daichi.html.
Jung, K., Kaiser, S., Böhm, S., Nieschulze, J., Kalko, E.K.V., 2012. Moving in three dimensions: effects of structural complexity on occurrence and activity of insectivorous bats in managed forest stands. J. Appl. Ecol. 49 (2), 523–531. https://doi.org/10.1111/j.1365-2664.2012.02116.x.
Lesak, A.A., Radeloff, V.C., Hawbaker, T.J., Pidgeon, A.M., Gobakken, T., Contrucci, K., 2011. Modeling forest songbird species richness using LiDAR-derived measures of forest structure. Remote Sens. Environ. 115 (11), 2823–2835. https://doi.org/10.1016/j.rse.2011.01.025.
Lovett, J.C., Wasser, S.K., 1993. Biogeography and Ecology of the Rainforests of Eastern Africa. Cambridge University Press, Cambridge.
Lumley, T., 2017. Thomas Lumley based on Fortran code by Alan Miller. leaps: Regression Subset Selection . R package version 3.0. https://CRAN.R-project.org/package=leaps.
MacArthur, R.H., MacArthur, J.W., 1961. On bird species diversity. Ecology 42 (3), 594–598. https://doi.org/10.2307/1932254.
Manzanera, J.A., García-Abril, A., Pascual, C., Tejera, R., Martín-Fernández, S., Tokola, T., Valbuena, R., 2016. Fusion of airborne LiDAR and multispectral sensors reveals synergic capabilities in forest structure characterization. GIScience Remote Sens. 53 (6), 723–738. https://doi.org/10.1080/15481603.2016.1231605.
Mbuthia, K.W., 2003. Ecological and Ethnobotanical Analyses for Forest Restoration in the Taita Hills, Kenya. Department of Botany. Miami University, Oxford, OH.
McGaughey, R.J., 2016. FUSION/LDV: Software for LIDAR data analysis and visualization. US Department of Agriculture, Forest Service, Pacific Northwest Research Station: Seattle, WA, USA (Version 3.60).
McRoberts, R.E., Winter, S., Chirici, G., Hauk, E., Pelz, D.R., Moser, W.K., Hatfield, M.A., 2008. Large-scale spatial patterns of forest structural diversity. Can. J. For. Res. 38 (3), 429–438. https://doi.org/10.1139/X07-154.
Miura, N., Jones, S.D., 2010. Characterizing forest ecological structure using pulse types and heights of airborne laser scanning. Remote Sens. Environ. 114 (5), 1069–1076. https://doi.org/10.1016/j.rse.2009.12.017.
Mononen, L., Auvinen, A.P., Packalen, P., Virkkala, R., Valbuena, R., Bohlin, I., Valkama, J., Vihervaara, P., 2018. Usability of citizen science observations together with airborne laser scanning data in determining the habitat preferences of forest birds. For. Ecol. Manage. 430, 498–508. https://doi.org/10.1016/j.foreco.2018.08.040.
Nölte, A., Meilby, H., Yousefpour, R., 2018. Multi-purpose forest management in the tropics: incorporating values of carbon, biodiversity and timber in managing Tectona grandis (teak) plantations in Costa Rica. For. Ecol. Manage. 422, 345–357. https://doi.org/10.1016/j.foreco.2018.04.036.
Omoro, L., Starr, M., Pellikka, P., 2013. Tree biomass and soil carbon stocks in indigenous forests in comparison to plantations of exotic species in the Taita Hills of Kenya. Silva Fennica 47 (2). https://doi.org/10.14214/sf.935.
Ørka, H.O., Gobakken, T., Næsset, E., Ene, L., Lien, V., 2012. Simultaneously acquired airborne laser scanning and multispectral imagery for individual tree species identification. Can. J. Remote Sens. 38 (2), 125–138. https://doi.org/10.5589/m12-021.
Pach, M., Podlaski, R., 2015. Tree diameter structural diversity in Central European forests with Abies alba and Fagus sylvatica: managed versus unmanaged forest stands. Ecol. Res. 30 (2), 367–384. https://doi.org/10.1007/s11284-014-1232-4.
Pellikka, P.K.E., Lötjönen, M., Siljander, M., Lens, L., 2009. Airborne remote sensing of spatiotemporal change (1955–2004) in indigenous and exotic forest cover in the Taita Hills, Kenya. Int. J. Appl. Earth Obs. Geoinf. 11 (4), 221–232. https://doi.org/10.1016/j.jag.2009.02.002.
Pellikka, P.K.E., Clark, B.J.F., Gosa, A.G., Himberg, N., Hurskainen, P., Maeda, E., Mwangombe, J., Omoro, L.M.A., Siljander, M., 2013. Chapter 13 – agricultural expansion and its consequences in the Taita Hills, Kenya. In: Paolo Paron, D.O.O., Christian Thine, O. (Eds.), Developments in Earth Surface Processes. Elsevier, pp. 165–179. https://doi.org/10.1016/B978-0-444-59559-1.00013-X.
Pellikka, P.K.E., Heikinheimo, V., Hietanen, J., Schäfer, E., Siljander, M., Heiskanen, J., 2018. Impact of land cover change on aboveground carbon stocks in Afromontane landscape in Kenya. Appl. Geogr. 94, 178–189. https://doi.org/10.1016/j.apgeog.2018.03.017.
Pfeifer, M., Gonsamo, A., Woodgate, W., Cayuela, L., Marshall, A.R., Ledo, A., Paine, T.C.E., Marchant, R., Burt, A., Calders, K., Courtney-Mustaphi, C., Cuni-Sanchez, A., Deere, N.J., Denu, D., de Tanago, J.G., Hayward, R., Lau, A., Macía, M.J., Olivier, P.I., Pellikka, P., Seki, H., Shirima, D., Trevithick, R., Wedeux, B., Wheeler, C., Munishi, P.K.T., Martin, T., Mustari, A., Platts, P.J.J.F.E., 2018. Tropical forest canopies and their relationships with climate and disturbance: results from a global dataset of consistent field-based measurements. Forest Ecosyst. 5 (1), 7. https://doi.org/10.1186/s40663-017-0118-7.
Phua, M.-H., Johari, S.A., Wong, O.C., Ioki, K., Mahali, M., Nilus, R., Coomes, D.A., Maycock, C.R., Hashim, M., 2017. Synergistic use of Landsat 8 OLI image and airborne LiDAR data for above-ground biomass estimation in tropical lowland rainforests. For. Ecol. Manage. 406, 163–171. https://doi.org/10.1016/j.foreco.2017.10.007.
Piiroinen, R., Fassnacht, F.E., Heiskanen, J., Maeda, E., Mack, B., Pellikka, P., 2018. Invasive tree species detection in the Eastern Arc Mountains biodiversity hotspot using one class classification. Remote Sens. Environ. 218, 119–131. https://doi.org/10.1016/j.rse.2018.09.018.
Piñeiro, G., Perelman, S., Guerschman, J.P., Paruelo, J.M., 2008. How to evaluate models: observed vs. predicted or predicted vs. observed? Ecol. Model. 216 (3), 316–322. https://doi.org/10.1016/j.ecolmodel.2008.05.006.
Potapov, P.V., Turubanova, S.A., Hansen, M.C., Adusei, B., Broich, M., Altstatt, A., Mane, L., Justice, C.O., 2012. Quantifying forest cover loss in Democratic Republic of the Congo, 2000–2010, with Landsat ETM+ data. Remote Sens. Environ. 122, 106–116. https://doi.org/10.1016/j.rse.2011.08.027.
Räsänen, M., Chung, M., Katurji, M., Pellikka, P., Rinne, J., Katul, G.G., 2018. Similarity in fog and rainfall intermittency. Geophys. Res. Lett. 45 (19). https://doi.org/10.1029/2018GL078837. 10,691-610,699.
RCoreTeam, 2017. R: A language and environment for statistical computing. In: R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the Great Plains with ERTS. In: Fraden, S.C., Marcanti, E.P., Becker, M.A. (Eds.), Third ERTS-1 Symposium. NASA, NASA SP-351, Washington D.C, pp. 309–317.
Schäfer, E., Heiskanen, J., Heikinheimo, V., Pellikka, P., 2016. Mapping tree species diversity of a tropical montane forest by unsupervised clustering of airborne imaging spectroscopy data. Ecol. Ind. 64, 49–58. https://doi.org/10.1016/j.ecolind.2015.12.026.
Schneider, F.D., Morsdorf, F., Schmid, B., Petchey, O.L., Hueni, A., Schimel, D.S., Schaepman, M.E., 2017. Mapping functional diversity from remotely sensed morphological and physiological forest traits. Nat. Commun. 8 (1), 1441. https://doi.org/10.1038/s41467-017-01530-3.
Stam, Å., Enroth, J., Malombe, I., Pellikka, P., Rikkinen, J., 2017. Experimental transplants reveal strong environmental effects on the growth of non-vascular epiphytes in Afromontane forests. Biotropica 49 (6), 862–870. https://doi.org/10.1111/btp.12472.
Thijs, K.W., Aerts, R., Musila, W., Siljander, M., Matthysen, E., Lens, L., Pellikka, P., Gulinck, H., Muys, B., 2014. Potential tree species extinction, colonization and recruitment in Afromontane forest relicts. Basic Appl. Ecol. 15 (4), 288–296. https://doi.org/10.1016/j.baae.2014.05.004.
Valbuena, R., Packalén, P., Martín-Fernández, S., Maltamo, M., 2012. Diversity and equitability ordering profiles applied to study forest structure. For. Ecol. Manage. 276, 185–195. https://doi.org/10.1016/j.foreco.2012.03.036.
Valbuena, R., Maltamo, M., Martín-Fernández, S., Packalen, P., Pascual, C., Nabuurs, G.-J., 2013a. Patterns of covariance between airborne laser scanning metrics and Lorenz curve descriptors of tree size inequality. Can. J. Remote Sens. 39 (sup1), S18–S31. https://doi.org/10.5589/m13-012.
Valbuena, R., Packalen, P., Mehtätalo, L., García-Abril, A., Maltamo, M., 2013b. Characterizing forest structural types and shelterwood dynamics from Lorenz-based indicators predicted by airborne laser scanning. Can. J. For. Res. 43 (11), 1063–1074. https://doi.org/10.1139/cjfr-2013-0147.
Valbuena, R., Vauhkonen, J., Packalen, P., Pitkänen, J., Maltamo, M., 2014. Comparison of airborne laser scanning methods for estimating forest structure indicators based on Lorenz curves. ISPRS J. Photogramm. Remote Sens. 95, 23–33. https://doi.org/10.1016/j.isprsjprs.2014.06.002.
Valbuena, R., Eerikäinen, K., Packalen, P., Maltamo, M., 2016. Gini coefficient predictions from airborne lidar remote sensing display the effect of management intensity on forest structure. Ecol. Indic. 60 (Supplement C), 574–585. https://doi.org/10.

1016/j.ecolind.2015.08.001.
Valbuena, R., Hernando, A., Manzanera, J.A., Martínez-Falero, E., García-Abril, A., Mola-Yudego, B., 2017b. Most similar neighbor imputation of forest attributes using metrics derived from combined airborne LIDAR and multispectral sensors. Int. J. Digital Earth 1–14. https://doi.org/10.1080/17538947.2017.1387183.
Valbuena, R., Hernando, A., Manzanera, J.A., Görgens, E.B., Almeida, D.R.A., Mauro, F., García-Abril, A., Coomes, D.A., 2017a. Enhancing of accuracy assessment for forest above-ground biomass estimates obtained from remote sensing via hypothesis testing and overfitting evaluation. Ecol. Model. 366, 15–26. https://doi.org/10.1016/j.ecolmodel.2017.10.009.
Valbuena, R., Maltamo, M., Mehtätalo, L., Packalen, P., 2017c. Key structural features of Boreal forests may be detected directly using L-moments from airborne lidar data. Remote Sens. Environ. 194, 437–446. https://doi.org/10.1016/j.rse.2016.10.024.
Valbuena, R., Hernando, A., Manzanera, J.A., Görgens, E.B., Almeida, D.R.A., Silva, C.A., García-Abril, A., 2019. Evaluating observed versus predicted forest biomass: R-squared, index of agreement or maximal information coefficient? Eur. J. Remote Sens. 52 (1), 1–14. https://doi.org/10.1080/22797254.2019.1605624.
Vierling, K.T., Vierling, L.A., Gould, W.A., Martinuzzi, S., Clawges, R.M., 2008. Lidar: shedding new light on habitat characterization and modeling. Front. Ecol. Environ. 6 (2), 90–98. https://doi.org/10.1890/070001.
Weiner, J., 1990. Asymmetric competition in plant populations. Trends Ecol. Evol. 5 (11), 360–364. https://doi.org/10.1016/0169-5347(90)90095-U.
Willson, M.F., 1974. Avian community organization and habitat structure. Ecology 55 (5), 1017–1029. https://doi.org/10.2307/1940352.
Zellweger, F., Roth, T., Bugmann, H., Bollmann, K., 2017. Beta diversity of plants, birds and butterflies is closely associated with climate and habitat structure. Glob. Ecol. Biogeogr. 26 (8), 898–906. https://doi.org/10.1111/geb.12598.
Zhang, Z., Cao, L., Mulverhill, C., Liu, H., Pang, Y., Li, Z., 2019. Prediction of diameter distributions with multimodal models using LiDAR data in subtropical planted forests. Forests 10 (2), 125. https://doi.org/10.3390/f10020125.
Zhu, Z., Wulder, M.A., Roy, D.P., Woodcock, C.E., Hansen, M.C., Radeloff, V.C., Healey, S.P., Schaaf, C., Hostert, P., Strobl, P., Pekel, J.-F., Lymburner, L., Pahlevan, N., Scambos, T.A., 2019. Benefits of the free and open Landsat data policy. Remote Sens. Environ. 224, 382–385. https://doi.org/10.1016/j.rse.2019.02.016.